Aspects of time series[*]

Attila Máté
Brooklyn College of the City University of New York

December 13, 2016

Contents

---

[*]Written for the course Mathematics 4506 (Time Series) at Brooklyn College of CUNY.

1

1 Introduction

In these notes we present some aspects of time series, mainly mathematical rather than statistical. There are many important mathematical issues that are often not discussed in time series text-books. One needs a basic understanding of complex Taylor series and the behavior of solutions of homogeneous linear difference equations to see the reason why certain models of stationary time series assume that the polynomials involved have all their zeros outside the unit circle. Frequency analysis demands some basic familiarity of Fourier series, the Fourier transform and trigonometric interpolation. This is provided without getting involved with convergence issues, even though occasionally we point out the presence of such issues. For example, at times we indicate that Riemann integration theory is inadequate to deal with certain of the subtleties, and one needs Lebesgue integration theory; we, however, try to keep such discussions non-technical, so as to make it available for advanced undergraduates. Wavelets are extremely important for time series, and, after a late start, they are exerting an increasing influence in applications for finance, yet introductory textbooks almost never discuss wavelets. Here we give the basic mathematical background, and not just calculational algorithms. The Kalman filter is ubiquitous in its application; we provide the basic mathematical background. An everyday cell phone uses wavelets for image representation, and it is running several Kalman filters. The notes are written with advanced undergraduates in mind, and issues of mathematical precision are often treated lightly. No applications are mentioned, and in general we underemphasized aspects of time series that are adequately represented in introductory texts. In particular, the book [11] gives an excellent coverage to these aspects. The book also gives numerous examples as to how to use the R programming language to build practical models of time series.

    There are many footnotes in these notes, to provide additional insight where including these comments in the main text would have interrupted the main flow of reasoning. In old times, printers used to complain about footnotes, since it was hard to typeset them and to make sure that the page had the correct size. This is no longer an issue with computerized typesetting, and the quantity of footnotes is simply a matter of writing style. These notes were written in LaTeX running under Debian Linux.

2 The multivariate normal distribution

2.1 The single variable normal distribution

Let σ and μ be reals, and assume that $\sigma > 0$. The random variable X is said to have a normal distribution with mean μ and standard deviation σ if its density function f_X is

$$(2.1) \qquad f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

The factor before the exponential ensures that

$$(2.2) \qquad \int_{-\infty}^{\infty} f_X(x)\,dx = 1.$$

This is easy to see, since

$$\left(\int_0^{\infty} e^{-x^2}\,dx\right)^2 = \int_0^{\infty} e^{-x^2}\,dx \int_0^{\infty} e^{-y^2}\,dy = \int_0^{\infty}\int_0^{\infty} e^{-x^2}e^{-y^2}\,dy\,dx$$

$$= \iint_{\{(x,y):\,x\geq 0,\ y\geq 0\}} e^{-x^2-y^2}\,dy\,dx = \iint_{\{(r,\theta):\,r\geq 0,\ 0\leq\theta\leq\pi/2\}} e^{-r^2} r\,dr\,d\theta.$$

The last integral was obtained by transforming the double integral in Cartesian coordinates to polar coordinates. This last integral is easily evaluated by iterated integration; it equals

$$\int_0^{\pi/2}\int_0^{\infty} e^{-r^2} r\,dr\,d\theta = \int_0^{\pi/2}\int_0^{\infty} e^{-t}\frac{1}{2}\,dt\,d\theta = \int_0^{\pi/2}\frac{1}{2}\,d\theta = \frac{\pi}{4};$$

the inner integral was evaluated by the substitution $t = r^2$. This implies that

$$(2.3) \qquad \int_{-\infty}^{\infty} e^{-x^2}\,dx = 2\int_0^{\infty} e^{-x^2}\,dx = 2\sqrt{\frac{\pi}{4}} = \sqrt{\pi}.$$

As we said, From here (2.2) follows by a simple change of variable. The single variable normal distribution with mean μ and standard deviation σ, i.e., variance σ^2 is denoted as $\mathcal{N}(\mu, \sigma^2)$.

2.2 The multivariate normal distribution

Writing A^T for the transpose of the matrix A, consider the random column vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)^T$. $\mathbf{X}$ is said to have a multivariate normal distribution if there is a random column vector $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_k)^T$ for some integer k with $0 \leq k \leq n$ whose components are independent random variables[2.1] each with distribution $\mathcal{N}(0,1)$, an $n \times k$ matrix A of reals, and an n-dimensional column vector $\boldsymbol{\mu}$ such that

$$(2.4) \qquad \mathbf{X} = A\mathbf{Z} + \boldsymbol{\mu}.$$

[2.1]What we mean here is that the whole collection of of these random variables is independent, which is a stronger condition than saying that they are *pairwise* independent; the latter means that any two of them are independent. In case we are given an infinite number of random variables, by their independence we mean that any finite subcollection is independent. We will always use independence in this sense; when we mean pairwise independence, we will explicitly say so.

If $k = n$ and A is a nonsingular matrix, then $\mathbf{X}$ is said to have a nondegenerate multivariate normal distribution; if $k < n$ or $k = n$ and A is a singular $n \times n$ matrix, then $\mathbf{X}$ is said to have a degenerate multivariate normal distribution.[2.2] While the degenerate case is important for statistics, since it can happen that the residuals in case of a least-squares fitting have a degenerate multivariate normal distribution, discussing the degenerate case is more complicated with the means at our disposal, since the joint density function does not exist in the degenerate case (this causes no difficulty with more advanced tools from measure theory).

2.3 The covariance matrix

Write $A = (a_{ij})_{1 \le i \le k\ 1 \le j \le k}$, and $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)^T$. Let p, q be integers with $1 \le p, j \le n$. We have $X_p = \sum_{i=1}^{k} a_{pi} Z_i + \mu_p$ and $X_q = \sum_{j=1}^{k} a_{qj} Z_j + \mu_q$. Let δ_{ij} be Kronecker's delta, that is

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \ne j. \end{cases}$$

Using the independence of the Z_i, we obtain

$$\mathrm{Cov}(X_p, X_q) = \mathrm{E}\big((X_p - \mu_p)(X_q - \mu_q)\big) = \mathrm{E}\Big(\sum_{i=1}^{k} a_{pi} Z_i \sum_{j=1}^{k} a_{qj} Z_j\Big)$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{k} a_{pi} a_{qj} \, \mathrm{E}(Z_i Z_j) = \sum_{i=1}^{k} \sum_{j=1}^{k} a_{pi} a_{qj} \delta_{ij} = \sum_{i=1}^{k} a_{pi} a_{qi}$$

Putting $\Sigma = A A^T$, the right-hand side is the entry in the pth row and the qth column of Σ. For this reason, Σ is called the covariance matrix of the random vector $\mathbf{X}$. For a random vector $\mathbf{X}$, we will write $\mathrm{Cov}(\mathbf{X})$ for its covariance matrix.[2.3] Note that if $\mathbf{X}$ is a random column vector and $E(\mathbf{X}) = \boldsymbol{\mu}$ then

(2.5) $$\mathrm{Cov}(\mathbf{X}) = \mathrm{E}\big((\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T\big).$$

Indeed, if $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ and $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)$, then

$$(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T$$

is an $n \times n$ matrix with the entry $(X_p - \mu_p)(X_q - \mu_q)$ in the pth row and qth column. Given a matrix (Z_{pq}) of random variables, its expectation is taken entry-wise, that is $\mathrm{E}\big((Z_{pq})\big) = \big(E(Z_{pq})\big)$.

2.4 The density function of a nondegenerate multivariate normal distribution

While the degenerate case is important for statistics, since it can happen that the residuals in case of a least-squares fitting have a degenerate multivariate normal distribution, discussing the degenerate case is more complicated with the means at our disposal, since the joint density function does not

---

[2.2]We may require A to be an $n \times n$ matrix. If A is an $n \times k$ matrix, we can replace it with the $n \times n$ matrix A' whose first k columns agree with those of A, and the remaining columns are zero.

[2.3]Note that there is no cause for confusion between the notation $\mathrm{Cov}(X, Y)$, with two arguments, denoting the covariance of two random variables and $\mathrm{Cov}(\mathbf{X})$, with a single argument, denoting the covariance matrix of the random vector $\mathbf{X}$.

exist in the degenerate case (this causes no difficulty with more advanced tools from measure theory). Hence, for discussing the density function, we assume that $k = n$ and the matrix A is nonsingular.

Writing $\mathbf{z} = (z_1,\, z_2,\, \ldots,\, z_k)^T$, the joint density function of $\mathbf{Z}$ is

$$(2.6) \qquad f_{\mathbf{Z}}(\mathbf{z}^T) = f_{\mathbf{Z}}(z_1, z_2, \ldots, z_n) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} z_i^2\right) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\mathbf{z}^T\mathbf{z}\right).$$

Assuming A is nonsingular, for $\mathbf{x} = A\mathbf{z} + \boldsymbol{\mu}$, writing we have $\mathbf{z} = A^{-1}(\mathbf{x} - \boldsymbol{\mu})$. The Jacobian matrix $\partial \mathbf{z}/\partial \mathbf{x}$ equals A^{-1}. Let $\Sigma = AA^T$ be the covariance matrix discussed above. Then we have $\det \Sigma = \det(AA^T) = \det A \det A^T = (\det A)^2$. Hence

$$(2.7) \qquad \left|\det \frac{\partial \mathbf{z}}{\partial \mathbf{x}}\right| = |\det(A^{-1})| = (\det \Sigma)^{-1/2}.$$

Furthermore

$$(2.8) \qquad \mathbf{z}^T\mathbf{z} = (\mathbf{x} - \boldsymbol{\mu})^T(A^{-1})^T A^{-1}(\mathbf{x} - \boldsymbol{\mu}) = (\mathbf{x} - \boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu});$$

the last equation holds since $\Sigma^{-1} = (AA^T)^{-1} = (A^T)^{-1}A^{-1}$. Note that $\mathbf{z}^T\mathbf{z} = \sum_{i=1}^{n} z_i^2 \geq 0$ unless $\mathbf{z} = \mathbf{0}$. Thus, the above equation with $\mathbf{y} = \mathbf{x} - \boldsymbol{\mu}$ shows that $\mathbf{y}\Sigma\mathbf{y} > 0$ unless $\mathbf{y} = \mathbf{0}$ (note that $\mathbf{y} = A^{-1}\mathbf{z} = \mathbf{0}$ only if $\mathbf{z} = \mathbf{0}$). Hence, the matrix Σ is positive definite – see [23, §35, pp. 159–]. We cannot recover the matrix A from Σ, there is, however, a unique positive definite matrix A' such that $(A')^2 = \Sigma$. We call A' as the square root of Σ. and we write $\sqrt{\Sigma}$ for this matrix, $\sqrt{\Sigma}$ is symmetric and it commutes with Σ – see [24, Subsection 9.5, p. 27]. For the density function $f_{\mathbf{X}}$ we have

$$f_{\mathbf{Z}}(z^T) \prod_{i=1}^{n} dz_i = f_{\mathbf{Z}}(\mathbf{z}^T) \left|\frac{\partial \mathbf{z}}{\partial \mathbf{x}}\right| \prod_{i=1}^{n} dx_i = f_{\mathbf{X}}(\mathbf{x}) \prod_{i=1}^{n} dx_i.$$

Thus, by equations (2.6) (2.7), and (2.8), we have

$$(2.9) \qquad f_{\mathbf{X}}(\mathbf{x}^T) = (2\pi)^{-n/2}(\det \Sigma)^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right).$$

According to this formula, if we take $A' = \sqrt{\Sigma}$, and $\mathbf{X}' = A\mathbf{Z} + \boldsymbol{\mu}$, then $\mathbf{X}$ and $\mathbf{X}'$ have exactly the same density functions, and so $\mathbf{X} = \mathbf{X}'$. So, in our considerations, we may, without loss of generality, assume that A is a positive definite symmetric matrix.[2.4]

2.5 Marginal distributions of the multivariate normal distribution

Given a positive definite symmetric matrix Σ, the matrix Σ^{-1} is also positive definite. This is because, given any nonzero column vector $\mathbf{x}$, with $A = \sqrt{\Sigma}$, put $\mathbf{u} = A^{-1}\mathbf{x}$, we have

$$\mathbf{x}^T\Sigma^{-1}\mathbf{x} = (A\mathbf{u})^T\Sigma^{-1}(A\mathbf{u}) = \mathbf{u}^T A^T\Sigma^{-1}A\mathbf{u} = \mathbf{u}^T\Sigma^{-1}AA\mathbf{u} = \mathbf{u}^T\Sigma^{-1}\Sigma\mathbf{u} = \mathbf{u}^T\mathbf{u} > 0;$$

For the third equation, note that A is symmetric and it commutes with Σ, and therefore also with Σ^{-1}; namely multiplying the equation $A\Sigma = \Sigma A$ by Σ^{-1} from both left and right, we obtain $\Sigma^{-1}A = A\Sigma^{-1}$.

[2.4]That is, it is positive definite if it is nonsingular. In the singular case we can only assume that it is positive semidefinite. We will comment on the case of singular A at the end.

Writing $\mathbf{y} = \mathbf{x} - \boldsymbol{\mu}$, $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T$, and $P = (p_{ij}) = \Sigma^{-1}$ we have

$$
\begin{aligned}
(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) = \mathbf{y}^T P \mathbf{y} &= \sum_{i=1}^n \sum_{j=1}^n y_i p_{ij} y_j \\
&= p_{11} y_1^2 + 2 y_1 \sum_{i=2}^n p_{1i} y_i + \sum_{i=2}^n \sum_{j=2}^n y_i p_{ij} y_j \\
&= p_{11} \left(y_1 + \sum_{i=2}^n \frac{p_{1i}}{p_{11}} y_i \right)^2 - \sum_{i=2}^n \frac{p_{1i}^2}{p_{11}} y_i^2 + \sum_{i=2}^n \sum_{j=2}^n y_i p_{ij} y_j \\
&= p_{11} \left(y_1 + \sum_{i=2}^n \frac{p_{1i}}{p_{11}} y_i \right)^2 + \sum_{i=2}^n \sum_{j=2}^n y_i \left(p_{ij} - \delta_{ij} \frac{p_{1i}^2}{p_{11}} \right) y_j
\end{aligned}
$$

(2.10)

where in the third equation we made use of the fact that $p_{ij} = p_{ji}$.

Substitute this into (2.9) and integrate with respect to dx_1. Noting that $dx_1 = dy_1$, we will use the substitution

$$
t = \frac{1}{\sqrt{2}} \sqrt{p_{11}} \left(y_1 + \sum_{i=2}^n \frac{p_{1i}}{p_{11}} y_i \right) ;
$$

observe that $p_{11} > 0$ since p is positive definite; – see [23, §35, pp. 159–]. Writing $\mathbf{y}_2 = (y_2, y_3, \ldots, y_n)^T$, equation (2.3) gives that the marginal density of $\mathbf{X}_2 = (X_2, X_3, \ldots, X_n)^T$ equals

$$
\begin{aligned}
f_{\mathbf{X}_2}(\mathbf{y}_2)^T) &= \int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{y})^T) \, dx_1 = (2\pi)^{-n/2} (\det \Sigma)^{-1/2} \sqrt{\pi} \sqrt{2} (p_{11})^{-1/2} \\
&\quad \exp \left(-\frac{1}{2} \sum_{i=2}^n \sum_{j=2}^n y_i \left(p_{ij} - \delta_{ij} \frac{p_{1i}^2}{p_{11}} \right) y_j \right) \\
&= (2\pi)^{-(n-1)/2} (\det \Sigma)^{-1/2} (p_{11})^{-1/2} \exp \left(-\frac{1}{2} \sum_{i=2}^n \sum_{j=2}^n y_i \left(p_{ij} - \delta_{ij} \frac{p_{1i}^2}{p_{11}} \right) y_j \right) .
\end{aligned}
$$

This has the form of the density function of a multivariate normal distribution. To make sure that it is indeed represents multivariate normal distribution, we only need to ascertain that the matrix with entries

$$
\left(p_{ij} - \delta_{ij} \frac{p_{1i}^2}{p_{11}} \right) \qquad (2 \le i, j \le n)
$$

is positive definite, i.e., that

$$
\sum_{i=2}^n \sum_{j=2}^n y_i \left(p_{ij} - \delta_{ij} \frac{p_{1i}^2}{p_{11}} \right) y_j > 0
$$

unless $y_2 = y_3 = \ldots = y_n = 0$. This is however immediate, since the the expression is identical to the right-hand side of (2.10) for the choice of y_1 that makes the expression under the square in the first term there zero, since we know that his right-hand side is positive unless $y_1 = y_2 = \ldots = y_n = 0$.

Once we know that the marginal density of $\mathbf{X}_2$ is a multivariate normal distribution, we can write this density function in a simpler form since the covariance matrix of $\mathbf{X}_2$ is the matrix obtained by deleting the first row and the first column of Σ (since the covariances of X_i and X_j are the entries of Σ).

2.6 The degenerate normal distribution

If the matrix A in equation (2.4) is singular, we are led to a degenerate normal distribution, and the discussion in Subsection 2.4 breaks down because the matrix Σ is singular; in this case Σ is only positive semidefinite (that is $\mathbf{x}\Sigma\mathbf{x} \geq 0$ always, but it can equal 0 even if $\mathbf{x} \neq \mathbf{0}$). However, if $\epsilon > 0$ and I is the $n \times n$ identity matrix, then the matrix $\Sigma + \epsilon I$ is nonsingular. This is because a matrix is singular if and only if 0 is one of its eigenvalues. As Σ is positive semidefinite, all its eigenvalues are nonnegative real numbers, and all the eigenvalue of $\Sigma + \epsilon I$ are of form $\lambda + \epsilon$ with λ and eigenvalue of Σ. The singular case then can be handled with replacing Σ by $\Sigma + \epsilon I$, and making $\epsilon \searrow 0$. The density function will exist for all $\epsilon > 0$, and so does the joint distribution function. The limit of the joint distribution when $\epsilon \searrow 0$ will define the joint distribution function for $\epsilon = 0$, while the joint density function will remain undefined. One needs some machinery from measure theory to define the density function (or, rather, more properly, the density measure) in case $\epsilon = 0$.

2.7 Independence and no correlation

We have the following

Theorem 2.1. *Assume the random variables X and Y have a joint normal distribution. If* $\mathrm{Cov}(X, Y) = 0$ *then X and Y are independent.*

Proof. Assume $\mathrm{Cov}(X, Y) = 0$; then the covariance matrix Σ of (X, Y) has form

$$\Sigma = \begin{pmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_Y^2 \end{pmatrix}.$$

If $\sigma_X = 0$ then X is constant with probability 1, in which case X and Y are independent; similarly, if $\sigma_Y = 0$. If $\sigma_X \neq 0$ and $\sigma_Y \neq 0$, then Σ is nonsingular, and we have

$$\Sigma^{-1} = \begin{pmatrix} \sigma_X^{-2} & 0 \\ 0 & \sigma_Y^{-2} \end{pmatrix}.$$

The joint density function of X and Y is given by

$$f_{X,Y}(x, y) = \frac{1}{2\pi \rho_X \rho_Y} \exp\left(-\frac{(x - \mu_1)^2}{2\sigma_X^2} - \frac{(y - \mu_2)}{2\sigma_Y^2} \right)$$

according to (2.9), and so $f_{X,Y}(x, y) = f_X(x)f_Y(y)$, since f_X and f_Y is given by (2.1) with appropriate modifications. $\qquad\square$

This result naturally extend to two vector variables $\mathbf{X}$ and $\mathbf{Y}$ having a joint normal distributions. If in the joint covariance matrix, each entry involving an $\mathbf{X}$ component and a $\mathbf{Y}$ component is zero,[2.5] then $\mathbf{X}$, and $\mathbf{Y}$ are independent as vector variables, that is that is, for joint the density function we have

$$f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = f_{\mathbf{X}}(\mathbf{x})f_{\mathbf{Y}}(\mathbf{y}).$$

The proof of this is similar to that of the above theorem; we omit the details.

2.8 Problem

Problem 2.1. Given a positive integer n, an $n \times n$ matrix A with real entries is called positive semidefinite if $\mathbf{x}^T A \mathbf{x} \geq 0$ for every n-dimensional column vector $\mathbf{x}$ with real components. Assume $\mathbf{X}$ is a random column vector with real entries. Show that its covariance matrix is positive semidefinite.

[2.5]In this case, one says that $\mathbf{X}$ and $\mathbf{Y}$ are *uncorrelated.*

3   Some background from complex function theory

The theory of complex functions of a single variable is an extension of single variable calculus, in that the functions are defined in a part of the complex plane, and the values are also complex numbers. Such a function f defined on a disk $D = \{z \in \mathbb{C} : |z - a| < r\}$, where a is a complex number, r is a positive real, and $\mathbb{C}$ denotes the set of complex numbers, is called differentiable in D if for any $z \in D$ the limit

$$\lim_{\zeta \to z} \frac{f(\zeta) - f(z)}{\zeta - z}$$

exists, and this limit is denoted as $f'(z)$ and is called the derivative of f at z. So far this is the same definition as given in real variable calculus, but the requirement for differentiability much more stringent. To explain this, note that is $z = x + iy$ where x and y are real, then $f(z)$ can be written as

$$f(z) = u(x, y) + iv(x, y),$$

where u and v are real functions of two real variables. The existence of the limit above means, in particular, that, for real h the limits

$$\lim_{h \to 0} \frac{f\big((x + h) + iy\big) - f(x + iy)}{h}$$

and

$$\lim_{h \to 0} \frac{f\big(x + i(y + h)\big) - f(x + iy)}{ih}$$

are equal. The equation of these limits can be written in terms of the functions u and v as

$$\frac{\partial}{\partial x} u(x, y) = \frac{\partial}{\partial y} v(x, y) \quad \text{and} \quad -\frac{\partial}{\partial y} u(x, y) = \frac{\partial}{\partial x} v(x, y).$$

These equations are called the Cauchy–Riemann equations. We will have no use for them in what follows, we mention them only to underline the differences between real and complex analysis.

The rules of differentiation (differentiation of sums, products, fractions, and composition of functions) are the same in complex variables as in real variables, but there are some features in complex analysis that are very different from what we know in real analysis. In particular, if f is differentiable in D then f' is also differentiable in D – nothing like this is true for real variables. Furthermore, if f is differentiable in D, then the Taylor series

$$\sum_{k=0}^{\infty} \frac{f(k)(a)}{k!} (z - a)^k$$

absolutely converges to f in D.

The above facts have the following consequence, important for time series: if $P(z)$ and $Q(z)$ are polynomials, and $Q(z)$ has no zeros in D, then the Taylor series of $P(z)/Q(z)$ at a absolutely converges in D. We will need this result with $a = 0$. If $P(z)$ and $Q(z)$ have real coefficients and a is real, then the Taylor series is $P(z)/Q(z)$ at a will have real coefficients, since the whole Taylor series can be determined by staying within the realm of real numbers. Determining the coefficients by repeated differentiation is usually to time-consuming, and it is easier to use polynomial division to do this in case $a = 0$, the main case of interest to us. The usual method of dividing polynomials can be used, but the polynomials need to be arranged in increasing powers, and the terms with the lowest power need to be divided at each step.

3.1 The natural exponential function with a complex argument

There are several ways to extend the natural exponential function $\exp x = e^x$ for complex values of x. One is to use the Taylor series

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!},$$

another one is using the limit

$$e^x = \lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n;$$

the latter approach has more intuitive appeal – see [25]. In the quoted note, one can find a proof of Euler's equation

$$(3.1) \qquad\qquad e^{ix} = \cos x + i \sin x,$$

where i is the imaginary unit. This equation is true for real and complex values of x; in [25], the proof given only for real x.[3.1]

4 Homogeneous linear recurrence equations

An equation of form

$$(4.1) \qquad\qquad \sum_{k=0}^{m} a_k y_{t-k} = 0 \qquad (a_0 \neq 0,\ a_m \neq 0,\ m > 0,\ -\infty < t < +\infty)$$

is called a *recurrence equation*, more precisely, a *homogeneous linear* recurrence equation. (If the right-hand side is replaced with some function of t that is not identically zero, then what he get is called an *inhomogeneous* recurrence equation. In this section, we will only discuss homogeneous recurrence equations.) Here a_k for integers k with $0 \leq k \leq m$ are given numbers, and we seek solutions y_n such that these equations are satisfied for all nonnegative integers n. m is called the order of this equation. The assumptions $a_0 \neq 0$ and $a_m \neq 0$ are reasonable in the sense that if either of these assumptions fail, the equation can be replaced with a lower order equation. It will be advantageous to work with complex numbers; i.e., the numbers a_k and y_n will be allowed to be complex. It is convenient to consider a solution of this equation as a vector

$$\mathbf{y} = \langle \ldots, y_{-2}, y_{-1}, y_0, y_1, y_2, \ldots \rangle$$

with two-way infinitely many components. These vectors can be added componentwise, that is

$$\langle \ldots, y_{-2}, y_{-1}, y_0, y_1, y_2, \ldots \rangle + \langle \ldots, z_{-2}, z-1, z_0, z_1, z_2, \ldots \rangle$$
$$= \langle \ldots, y_{-2} + z_{-2}, y_{-1} + z_{-1}, y_0 + z_0, y_1 + z_1, y_2 + z_2, \ldots \rangle,$$

and can be multiplied by scalars, that is

$$\alpha \langle \ldots, y_{-2}, y_{-1}, y_0, y_1, y_2, \ldots \rangle = \langle \ldots, \alpha y_{-2}, \alpha y_{-1}, \alpha y_0, \alpha y_1, \alpha y_2, \ldots \rangle.$$

[3.1]Euler's equation has an appealing intuitive content for real x if one considers the extension of $\exp x$ for complex x using the limit above. The equation is easy to prove for complex x if one uses the Taylor series of $\exp x$, $\cos x$, and $\sin x$, but such a proof has no intuitive content.

The solution vectors form an m-dimensional vector space. First, they form a vector space, since if $\mathbf{y}$ and $\mathbf{z}$ are solutions then $\alpha \mathbf{y} + \beta \mathbf{z}$ is also a solution. It is also clear that the dimension of this vector space is m since each solution is determined if we specify the number y_i for m consecutive integer, for example, for each each integer i with $0 \leq i \leq m - 1$ (indeed, y_t for $j \geq m$ is then determined by the recurrence equation, as $a_0 \neq 0$ and $a_m \neq 0$), and the numbers y_i for these m consecutive integers can be specified arbitrarily.

Write

$$(4.2) \qquad P(\zeta) = \sum_{k=0}^{m} a_k \zeta^k.$$

The polynomial $P(\zeta)$ is called the characteristic polynomial of the recurrence equation (4.1), and the polynomial equation $P(\zeta) = 0$ is called its characteristic equation. Here ζ is a complex variable.[4.1]

4.1 The forward and backward shift operators

The backward shift operator B on functions of defined on the set of all integers $\mathbb{Z}$ is given by writing $Bf(t) = f(t-1)$;[4.2] The powers of the operator B can be defined by $B^n f(t) = B\big(B^{n-1}f\big)f(t)\big)$ in addition, we can also use the identity operator I. Polynomials of the operator B will be called difference operators.[4.3] y_t will be considered as a function of n, and the operator B on y_t will act according to the equation $By_t = y_{t-1}$.[4.4] The recurrence equation (4.1) can be written in terms of the operator B as

$$(4.3) \qquad \left(\sum_{k=0}^{m} a_k B^k\right) y_t = 0 \qquad (t \in \mathbb{Z}).$$

The forward shift operator on function defined on the set of all integers $\mathbb{Z}$ given by writing $Ef(t) = f(t+1)$. We have $B = E^{-1}$, so equation (4.3) can also be written as $P(E^{-1})y_t = 0$. Multiplying both sides by E^m (the degree of $P(\zeta)$) makes no difference, since this equation is supposed to hold for all $t \in Z$, so this equation is more conveniently written as

$$E^m P(E^{-1})y_t = 0.$$

Observe that $Q(\zeta) \stackrel{def}{=} \zeta^m P(\zeta^{-1})$ is also polynomial, and this is called the characteristic polynomial of equation (4.1) when the equation is written in terms of the forward shift operator. When discussing difference equations, usually the forward shift operator is used, but in the theory of time series it is more common to use the backward shift operator. The solutions of equation (4.2) will be discussed in terms of the zeros of

---

[4.1]Or an indeterminate, from an alternative viewpoint. An indeterminate is a symbolic variable used in defining a polynomial ring, and is not to be interpreted as representing a number.

[4.2]The backward shift operator is always associated with a variable; if more than one variable were associated with backward shift operators, the notation should indicate the variable in question as well, for example E_t would shift the variable t forward, while E_s would shift the variable s, etc.

In a more rigorous treatment, however, B always acts on the function, and not the variable. That is, Bf is the function such that $(Bf)(t) = f(t-1)$ for all t. It is, however, useful to maintain the fiction that B acts on a variable in order not to complicate the notation too much.

[4.3]A basic difference operation is the backward difference operator $\nabla = I - B$. Since we have $B = I - \nabla$, a recurrence equation can also be written in terms of the backward difference operator. For this reason, a recurrence equation is also called a *difference equation*.

[4.4]It would be formally more correct, but less convenient, to say that B acts on vectors $\mathbf{y}\langle \ldots, y_{-2}, y_{-1}, y_0, y_1, y_2, \ldots \rangle$. In fact, properly, the vector $\mathbf{y}$ can be considered a function on the set of integers $\mathbb{Z}$, where y_t stands for $\mathbf{y}(t)$.

the equation $P(\zeta) = 0$. Since $P(\rho) = 0$ if and only if $Q(1/\rho) = 0$.[4.5] Hence, results stated in terms of the zeros of $P(\zeta)$ can easily be also described in terms of the zeros of $Q(\zeta)$; in fact, the latter description is more common – except when discussing time series.

By solving the characteristic equation, the characteristic polynomial can be factored as the product of m linear factors; assuming that λ_j is a zero[4.6] of multiplicity m_j of the characteristic polynomial for j with $1 \leq j \leq N$ (the λ_j's are assumed to be pairwise distinct), we have

$$\sum_{k=0}^{m} a_k \zeta^k = a_m \prod_{j=1}^{N} (\zeta - \lambda_j)^{m_j}, \qquad \text{where} \quad \sum_{j=1}^{N} m_j = m;$$

the second equation here just says that the above polynomial equation (of degree m) has m roots, counting multiplicities. The difference operator in recurrence equation (4.3) has a corresponding factorization:

$$\sum_{k=0}^{m} a_k B^k = a_m \prod_{j=1}^{N} (B - \lambda_j)^{m_j};$$

here $B - \lambda_j$ could also have been written as $B - \lambda_j I$, but the identity operator is often omitted when is has a number coefficient. This is because the rules of algebra involving polynomials of the variable ζ and polynomials of the forward shift operator B are the same.[4.7]

The degree of a polynomial $P(t)$ of t will be denoted by $\deg P(t)$; the constant polynomial that is not identically zero will have degree zero, and the identically zero polynomial will have degree -1. Then we have

Lemma 4.1. *Let λ and η be nonzero complex numbers, and let $P(t)$ be a polynomial of t that is not identically zero. Then*

$$(B - \lambda)P(t)\eta^{-t} = Q(t)\eta^{-t},$$

where $Q(t)$ is another polynomial of t such that $\deg Q(t) = \deg P(t)$ if $\lambda \neq \eta$ and $\deg Q(t) = \deg P(t) - 1$ if $\lambda = \eta$.

Proof. Given an integer $k \geq 0$, we have

$$(B - \lambda)t^k \eta^{-t} = (t-1)^k \eta^{-(t-1)} - \lambda t^k \eta^{-t} = \sum_{j=0}^{k} \binom{k}{j} t^j (-1)^{k-j} \eta^{-t+1} - \lambda t^k \eta^{-t}$$

$$= \left((\eta - \lambda)t^k + \eta \sum_{j=0}^{k-1} \binom{k}{j} t^j (-1)^{k-j} \right) \eta^{-t};$$

the second equality was obtained by using the Binomial Theorem. This equation says it all; if $\lambda = \eta$ then the term involving t^k *will cancel*, and if $\lambda \neq \eta$ then this term *will not cancel*. In the former case, the operator lowers the degree of t^k in the term $t^k \eta^{-t}$ by one. (In this case, if t^k is the term of the highest degree of the polynomial $P(t)$, then the resulting term $-\eta\binom{k}{1}t^{k-1}\eta^{-t}$ will not cancel against the terms resulting from lower degree terms of $P(t)$, since the degrees of those terms will also be lowered.) The proof is complete. $\qquad\square$

[4.5] $\rho = 0$ cannot happen here, since $P(0) = a_0 \neq 0$ by our assumptions. Similarly, $Q(0) = a_m \neq 0$.

[4.6] A *zero* of a polynomial is a *root* of the equation obtained by equating the polynomial to zero.

[4.7] In particular, given complex numbers λ and η the operators $B - \lambda$ and $B - \eta$ commute; that is

$$(B - \lambda)(B - \eta) = (B - \eta)(B - \lambda).$$

Note that B does not commute with expressions involving t. For example, $tBt^2 = t(t-1)^2$, and $t^2 Bt = t^2(t-1)$.

4.2 Linear independence of certain functions

Functions here mean functions on $\mathbb{Z}$; instead of the word "function" we could have used the phrase "two-way infinite sequence." The lemma just established has several important corollaries.

Corollary 4.1 (Linear Independence). *Let $r \geq 1$ be an integer. Let $f_k(t) = P_k(t)\lambda_k^{-t}$ be functions of t for k with $1 \leq k \leq r$, where $P_k(t)$ is a polynomial of t that is not identically zero, and λ_k is a nonzero complex number, such that if $1 \leq k < l \leq r$ then either $\lambda_k \neq \lambda_l$, or if $\lambda_k = \lambda_l$ then $\deg P_k(t) \neq \deg P_l(t)$. Then the functions f_k are linearly independent.*

Proof. Assume, on the contrary, that we have

$$\sum_{k=1}^{r} c_k P_k(t)\lambda_k^{-t} \equiv 0,$$

where not all the complex coefficients c_k are zero ($\equiv$ here means that equality holds identically; in the present case this means that equality holds for every integer t). We will show that this equation cannot hold. To this end, without loss of generality, we may assume that none of the coefficients are zero, since the terms with zero coefficients can simply be discarded. Further, we may assume that among the terms $P_k(t)\lambda_k^{-t}$ the polynomial $P_1(t)$ is the one that has the highest degree (other polynomials $P_k(t)$ with nonzero c_k for $\lambda_k \neq \lambda_1$ may have the same degree, but not higher). Let d be the degree of $P_1(t)$. Then

$$(B - \lambda_1)^d \Big(\prod_{\substack{k:2\leq k\leq r, \\ \lambda_k \neq \lambda_1}} (B - \lambda_k)^{d+1} \Big) \sum_{k=1}^{r} c_k P_k(t)\lambda_k^{-t} = c\lambda_1^{-t},$$

with a nonzero c. The product is taken for all k for which λ_k is different from λ_1.[4.8] The reason for this equation is that the difference operator $(B - \lambda_k)^{d+1}$ annihilates the term $P_k(t)\lambda_k^{-t}$ when $\lambda_k \neq \lambda_1$ according to the Lemma 4.1 above, (since $\deg P_k(t) \leq d$). These operators will not change the degree of the polynomial in the term $P_1(t)\lambda_1^{-t}$ according to the same Lemma (because $\lambda_k \neq \lambda_1$). The operator $(B-\lambda_1)^d$ will annihilate the term $P_k(t)\lambda_k^{-t}$ in case $\lambda_k = \lambda_1$ and $k \neq 1$ (since $\deg P_k(t) < d$ in this case, according to our assumptions). Finally, the operator $(B - \lambda_1)^d$ lowers the degree of $P_1(t)$ by d in the term $P_1(t)\lambda_1^{-t}$ according to the Lemma (while none of the other operators change the degree of $P_1(t)$ in this term, as we mentioned). Hence, after the application of the above difference operators, the resulting function will be $c\lambda_1^{-t}$ with $c \neq 0$; this confirms the above equation. So, applying the difference operator to both sides of the equation expressing linear dependency, we obtain that

$$c\lambda_1^{-t} \equiv 0,$$

while $c \neq 0$. This is a contradiction since $\lambda_1 \neq 0$ according to assumptions, showing that the functions in question are linearly independent. □

4.3 The solution of the recurrence equation

Corollary 4.2 (Solution of the Homogeneous Equation). *Assuming*

$$\sum_{k=0}^{m} a_k \zeta^k = a_m \prod_{j=1}^{N} (\zeta - \lambda_j)^{m_j}, \qquad where \quad \sum_{j=1}^{N} m_j = m,$$

[4.8] This arrangement is of course highly redundant, because if $\lambda_k = \lambda_l$, there is no need to take both of the factors $(B - \lambda_k)^{d+1}$ and $(B - \lambda_l)^{d+1}$, but such redundancy is harmless and it serves to simplify the notation.

and the λ_j's are pairwise distinct, the functions $t^r \lambda_j^{-t}$ for r and j with $0 \le r < m_j$ and $1 \le j \le N$ represent m linearly independent solutions of the difference equation

$$\left(\sum_{k=0}^{m} a_k B^k\right) y_t = 0.$$

Proof. The linear independence of the functions claimed to be representing the solutions have been established in Corollary 4.1. Since a recurrence equation of order m can have at most m linearly independent solutions, these functions will represent a complete set of linearly independent solutions. To see that each of these functions is a solution, it is enough to note according to the equation

$$\sum_{k=0}^{m} a_k B^k = a_n \prod_{j=1}^{N} (B - \lambda_j)^{m_j}$$

that, in view of Lemma 4.1, the difference operator

$$(B - \lambda_j)^{m_j}$$

annihilates the function $t^r \lambda_j^{-t}$ for $r < m_j$. $\qquad\square$

Thus we exhibited m linearly independent solutions of equation (4.1). If follows that any solution of (4.1) is a linear combination of these solutions.

4.4 The inhomogeneous linear recurrence equation

Given b_t for all $t \in \mathbb{Z}$, the equation

(4.4)
$$\sum_{k=0}^{m} a_k y_{t-k} = b_t \qquad (t \in \mathbb{Z})$$

is called an inhomogeneous recurrence equation, with (4.1) as the corresponding homogeneous equation. If the vectors

$$\mathbf{y}_1 = \langle \dots, y_{-2}^{(1)}, y_{-1}^{(1)}, y_0^{(1)}, y_1^{(1)}, y_2^{(1)}, \dots \rangle \quad \text{and} \quad \mathbf{y}_2 = \langle \dots, y_{-2}^{(2)}, y_{-1}^{(2)}, y_0^{(2)}, y_1^{(2)}, y_2^{(2)}, \dots \rangle$$

are solutions of the inhomogeneous equation, then, clearly, $\mathbf{y}_2 - \mathbf{y}_1$ is a solution of the homogeneous equation. Stated in another way, if we find a solution $\mathbf{y}_p$ of the inhomogeneous equation, then every solution of the inhomogeneous equation can be obtain as $\mathbf{y}_p + \mathbf{y}_h$, where $\mathbf{y}_h$ is a solution of the homogeneous equation. The solution y_p is often called a *particular* solution.

4.5 Problems

Problem 4.1. The Fibonacci numbers y_t, $t = 0, 1, 2, \dots$ are defined by the equations $y_0 = 0$, $y_1 = 1$ and $y_t = y_{t-1} + y_{t-2}$ for every integer $t \ge 2$. Write a formula expressing y_t.

Problem 4.2. Write a difference operator that annihilates all but the first term in the expression

$$c_1 t^3 \cdot 3^{-t} + c_2 t^4 \cdot 2^{-t} + c_3 t^2 \cdot 5^{-t},$$

while it reduces the first term to $c \cdot 3^{-t}$, where c is a nonzero constant (it is assumed that $c_1 \ne 0$).

5 Differencing and other transformations of time series

5.1 Stationary time series

A sequence
$$\{Y_t\} = \langle Y_t : t \in \mathbb{Z} \rangle$$
of random variables Y_t is called a time series. Considering two-way infinite time series is an idealization. A time series is usually observed in a finite interval. $\{Y_t\}$ is called *strictly stationary* if given any $n \geq 0$, the joint distribution of the sequence of random variables $\langle Y_{t+k} : 0 \leq k \leq n \rangle$ does not depend on $t \in Z$. It is called *stationary* if $\mathrm{E}(|Y_t|^2) < \infty$ for all t, $\mathrm{E}(Y_t)$ does not depend on t,[5.1] and for any $n \in \mathbb{Z}$, the covariance $\mathrm{Cov}(Y_t, Y_{t+n})$ does not depend on t.[5.2]

5.2 Time series and recurrence equations

Let $P(x)$ be a polynomial with constant term 1, and assume the time series $\{Y_t\}$ satisfies the equation

(5.1) $$P(B)Y_t = E_t \qquad (t \geq 0),$$

where E_t is the error at time t; at this point, we do not assume anything about E_t;[5.3] in fact, we would treat the whole question as involving a numerical series, except that our concern is to turn the time series $\{Y_t\}$ into a stationary series. Normally, equations of the type (5.1) are considered in autoregressive models of time series. Here, we are not concerned with modeling; in fact, we are not assuming that our time series is stationary, and autoregressive modeling is usually considered for stationary time series. So, before modeling, one wants to turn the time series into a stationary time series. The main tools for this is differencing and seasonal differencing, and other transformations.

One might ask, why would a time series satisfy an equation such as (5.1). In fact, Section 23, especially Subsection 23.2 gives an answer. State space models describe how the random variable Y_t produced at time t is produced by the state of the system. Such states may be natural for all time series; however, in most situations, not much if anything can be known about the state. Only in models of engineering processes would be a more or less clear understanding of states. Usually, the only choice one has is to try to model the time series, whatever mechanism produces it.

We do not assume that the polynomial $P(x)$ is known to the person analyzing the time series $\{Y_t\}$; in fact, we assume that it is not known. The only use we are making of equation (5.1) is to explain certain patterns of behavior of the time series that is observed by analyst, without knowing anything about this equation. All the actions described below to remedy the undesirable patterns of behavior are to be taken without any knowledge of this equation. On the other hand, the effects of these actions can be best explained with this equation in sight.

5.3 Differencing

Differencing means applying the operator $\nabla \overset{def}{=} I - B$ to Y_t, and considering the time series $\{\nabla Y_t\}$, and considering what equation the latter time series satisfies. We can analyze the effects of such a transformation in terms Lemma 4.1 and Corollary 4.2.

[5.1] The assumption $\mathrm{E}(|Y_t|^2) < \infty$ implies that $\mathrm{E}(Y_t)$ exists.

[5.2] Note that the assumption $\mathrm{E}(|Y_t|^2) < \infty$ implies that $|\mathrm{E}(Y_t)| < \infty$, we have

$$|\mathrm{E}(X)| \leq \mathrm{E}(|X|) = \mathrm{E}(|X|) \cdot \mathrm{E}(1) \leq \mathrm{E}(|X|^2) \cdot \mathrm{E}(1^2) = \mathrm{E}(X^2),$$

where the second inequality holds by the Schwarz inequality (see 5.2).

[5.3] If the degree of $P(x) = m$, then Y_t needs to be defined for all $t \geq -m$ for the above equation to make sense.

First note that equation (5.1) is an inhomogeneous recurrence equation, and, according to Subsection 4.4 its solution is a particular solution of this equation and a solution of the homogeneous equation. In Corollary 4.2 we described the basic solution of the homogeneous equation; the general solution of the homogeneous equation is a linear combination of these basic solutions. The coefficients of this linear combination are determined by the initial conditions, i.e., the initial observations of the time series $\{Y_t\}$. Because of the random nature of these observations, all basic solutions of the homogeneous equation are likely to occur with nonzero coefficients.

According to Corollary 4.2, the solutions of the homogeneous equation involve terms of form $t^k \lambda^{-t}$, where λ is a zero of the polynomial $P(x)$ with multiplicity greater than k ($k \geq 0$). If $|\lambda| > 1$ then $\lim_{t \to \infty} t^k \lambda^{-t} = 0$, so such terms cause no trouble in the long run, i.e., they do not prevent the time series from being stationary, at least asymptotically. On the other hand, if $|\lambda| < 1$ or if $|\lambda| = 1$ and $k \geq 1$ $\lim_{t \to \infty} t^k \lambda^{-t} = \infty$, so in this case the time series $\{Y_t\}$ will not be stationary. In view of Lemma 4.1, the operator $\nabla = -(I - B)$ has essentially no effect on the term $t^k \lambda^{-t}$, more precisely, it will not change its degree k unless $\lambda = 1$. So differencing is of no use unless $P(1) = 0$.

On the other hand, if 1 is a p-fold zero of $P(x)$, then p successive differencing will help. This can be seen as follows. In this case, we have $P(x) = R(x)(x-1)^p$, where $R(x)$ is a polynomial such that $R(1) \neq 0$. Then we have $P(x) = (-1)^p R(x)(1-x)^p$ and so $P(B) = (-1)^p R(B) \nabla$, and equation (5.1) can also be written as

$$R(B)\nabla^p Y_t = (-1)^p E_t \qquad (t \geq 0).$$

This is an equation for $\nabla^p Y_t$ where the characteristic polynomial $R(x)$ no longer has a zero at 1. But the other zeros have not been dealt with, and the remaining zeros λ with $|\lambda| < 1$ will cause trouble.

If λ is a zero of $P(x)$ with $|\lambda| > 1$, the terms involving λ^{-t} in the solution of the homogeneous equation associated with equation (5.1) will tend to ∞ in absolute value as $t \to -\infty$; so how come we are not concerned about these zeros. One answer might be that we are concerned about the future of the time series, and not its past; but there is another answer. What ever happened in the past, the errors E_t in equation (5.1) were such that they accommodated whatever values the time series assumed in the past. So the coefficients of the various terms involving λ^{-t} were such that Y_t remained within certain bounds in the past (if it indeed did). On the other hand, we have no such control over the future, especially since the future errors E_t are random, so we very much need to be concerned with the troublesome terms λ^{-t} with $|\lambda| < 1$, since these term will tend to ∞ when $t \to \infty$.

5.3.1 Inverting differencing (integrating)

Having obtained the time series $\{X_t\}$ by differencing $\{Y_t\}$, we build a model for model for $\{X_t\}$. Then we can apply this model to analyze $\{Y_t\}$ by restoring it from the modeled time series. Assuming

(5.2) $$X_t = \nabla Y_t = Y_t - Y_{t-1} \qquad (t \in \mathbb{Z}),$$

we have

$$Y_t = Y_{t-1} + X_t,$$

so, given the sequence $\{X_t\}$ and an initial value for $\{Y_t\}$, we can easily restore the whole sequence. For example

$$Y_t = Y_0 + \sum_{k=1}^{t} X_k \qquad (t > 0).$$

Given Y_0, restoration for Y_t for $t < 0$ is also possible, given equation (5.2), but usually is of no interest in the context of time series. If we have performed several differencing, we need to perform inverting the same number of times.

5.4 Seasonal differencing

If the time series $\{Y_t\}$ shows a periodic behavior of a period d for some d, such a periodicity is called seasonality. The origin of the term is that certain time series sampled once a month often naturally show seasonality of period 12, since often such time series are affected by the seasons of the year. In this case, one usually applies the seasonal differencing operator $\nabla_d \stackrel{def}{=} I - B^d$. The effect of this can also be analyzed in terms of equation (5.1). In terms of solutions of the homogeneous equation corresponding to this equation can be explained by the presence of a term λ^{-t} among the solutions of the of the homogeneous equation that is periodic with period d; this in possible only if $\lambda = e^{2\pi i l/d}$ for some integer l. This means that $P(e^{2\pi i l/d}) = 0$. For simplicity, assume that $e^{2\pi i l/d}$ is a simple zero (a zero of multiplicity 1) of $P(x)$. In fact, assume all the terms λ^{-t} for $\lambda = e^{2\pi i l/d}$ for any integer l with $1 \leq l < d$ causing periodicity of period d are present. Then $P(e^{2\pi i l/d}) = 0$ for l with $1 \leq l < d$. The zeros of the equation $x^d - 1 = 0$ are $e^{2\pi i l/d}$ for l with $0 \leq l < d$. We have

$$x^d - 1 = \prod_{l=0}^{d-1}(x - e^{2\pi i l/d}),$$

and so, putting

(5.3) $$Q(x) = \prod_{l=1}^{d-1}(x - e^{2\pi i l/d}) = \frac{x^d - 1}{x - 1} = \sum_{l=0}^{d-1} x^l,$$

the polynomial $Q(x)$ must be a divisor of $P(x)$; i.e., $P(x) = R(x)Q(x)$ holds for some polynomial $R(x)$. Assuming, for the sake of simplicity, that each $e^{2\pi i l/d}$ $(1 \leq l < d)$ is a simple zero (i.e., a zero of multiplicity 1) of $P(x)$, the numbers $e^{2\pi i l/d}$ are no longer zeros of $R(x)$. Now, equation (5.1) can be written as

$$R(B)\ Q(B)Y_t = E_t \qquad (t \geq 0).$$

This is an inhomogeneous equation for $Q(B)Y_t$. The corresponding homogeneous equation no longer has the seasonality terms $e^{2\pi i l/d}$ $(1 \leq l < d)$. Thus, considering

$$Q(B)Y_t = \sum_{l=0}^{d-1} B^l Y_t$$

instead of Y_t, we successfully removed seasonality. If $e^{il/d}$ is a multiple zero of $P(x)$, then we have to repeat this process in order to remove seasonality.

However, often this is not what is done in practice. One takes

$$\nabla_d Y_t = (I - B^d)Y_t = (I - B)\ Q(B)Y_t = \nabla\ Q(B)Y_t.$$

The time series $Q(B)Y_t$ no longer has seasonality. Assuming that $Q(B)Y_t$ is stationary, the differencing with ∇ on the left is unnecessary, and in our opinion it should not be done, since it amounts to overdifferencing; that is, applying a difference operator to a time series where such application is not necessary. The paper [9], or a shorter blog [10] by the same author, discusses the danger

of overdifferencing. The article [2] notes problem of the overdifferencing with seasonal models, and compares the overdifferenced model with another model that avoids overdifferencing, but it does not seem to state the simple mathematical reason that causes overdifferencing in our opinion. The lecture note [33, p. 6 of lec4-08.pdf] also points out that the polynomial $Q(B)$ rather than $1 - B^d$ should be used to remove seasonality.

5.4.1 Inverting seasonal differencing

Having obtained the time series $\{X_t\}$ by seasonal differencing and building a model for it, we want this model adapted for the original time series. Assuming

$$X_t = Q(B)Y_t \qquad (t \in \mathbb{Z}),$$

with $Q(x)$ given in equation (5.3), we have

(5.4) $$(I - B)X_t = (I - B^d)Y_t.$$

That is, writing $Z_t = (I - B)X_t = \nabla X_t$, we have

$$Z_t = Y_t - Y_{t-d}.$$

That is, if Y_m is given for m with $0 \leq m < d$ and the time series $\{X_t\}$, we can restore the time series $\{Y_t\}$. Indeed, given X_t for all $t \in \mathbb{Z}$, we can calculate Z_t, and then, for $n > 0$ and

$$Y_{m+nd} = Y_m + \sum_{k=1}^{n} Z_{m+kd} \qquad (t > 0).$$

If we have done several seasonal differencing, we need to repeat above steps step of inverting seasonal differencing.

In equation (5.4), we did a differencing by ∇, and, as the right-hand side shows, this amounts to calculate $\nabla_d Y_t$, in spite of having said above that this may amount to overdifferencing; this observation, however, misses the main point. We model the time series $\{X_t\}$, and we use this model to build a model for $\{Y_t\}$. That is, we calculate ∇X_t only after we modeled X_t, and calculating it is only used as a step to express Y_t in terms of X_t. The problem with overdifferencing is that it tries to build a model for ∇X_t instead of building it for X_t.

5.5 Logarithmic and other transformations

If the polynomial $P(x)$ in equation (5.1) has a zero λ with $|\lambda| < 1$ then, as we pointed out above, this zero will cause trouble, and differencing or seasonal differencing will not help. In this case, one might consider a logarithmic transformation, that is, instead of $\{Y_t\}$ one might one to study the time series $\{\log Y_t\}$.[5.4], assuming $Y_t > 0$ (if not, one might take $\log(cY_t + d)$ with an appropriate constants c and d. There may be other reasons to consider a logarithmic transformation. For example, in stock prices, one is usually concerned with percentage gains, i.e., multiplicative gains, and taking logarithms converts these to additive gains, which are technically easier to handle. Other transformations one may consider is to take

$$X_t = \frac{Y_t^\lambda - 1}{\lambda}$$

[5.4]We use $\log x$ to denote the natural logarithm of x. This is common mathematical practice, and $\ln x$ is rarely used in mathematical writing.

for some fixed $\lambda > 0$. if $Y_t > 0$. Incidentally, note that

$$\lim_{\lambda \searrow 0} \frac{x^\lambda - 1}{\lambda} = \log x$$

for all $x > 0$, as one can easily verify by l'Hospital's rule.

5.6 Convolutions and linear filters

Given two functions f and g on $\mathbb{R}$, their convolution $f * g$ is defined as

(5.5)
$$(f * g)(x) \stackrel{def}{=} \int_{-\infty}^{\infty} f(x - t)g(t)\, dt = \int_{-\infty}^{\infty} f(\tau)g(x - \tau)\, d\tau$$

provided the integral on the right exists; the second equation is obtained by the substitution $\tau = x - t$.[5.5] For two two-way infinite sequences (i.e., functions on $\mathbb{Z}$) we define[5.6]

(5.6)
$$(f * g)(n) \stackrel{def}{=} \sum_{k=-\infty}^{\infty} f(n - k)g(k) = \sum_{l=-\infty}^{\infty} f(l)g(n - l).$$

These equations show that convolution is a commutative operation, that is, $f * g = g * f$ both for functions and for sequences.

If one thinks of two-way infinite sequences as functions on $\mathbb{Z}$, then one can think of a (one-way) infinite sequence as a function f on $\mathbb{Z}^+$, the set of positive integers. Then a subsequence g of f can be thought of as the function $f \circ h$, where $h : \mathbb{Z}^+ \to \mathbb{Z}^+$ is a strictly increasing function.

In time series analysis, a convolution is usually called a linear filter. If $\{Y_t\}$ is a time series, then one can take a (usually fixed) number sequence $\{h_t\}$, and define the filtered time series $\{X_t\}$ as the convolution

(5.7)
$$X_t = \sum_{k=-\infty}^{\infty} Y_{t-k}h_k.$$

If one wants to analyze the time series $\{Y_t\}$ in real time, then one also needs to assume that the filter has no future dependence, that is, $h_k = 0$ for $k < 0$.

5.6.1 Moving average

Given a positive integer n, the following filter is called a *moving average* filter of length n: in equation (5.7) put

$$h_t = \begin{cases} 1/n & \text{if } 0 \le t < n, \\ 0 & \text{otherwise.} \end{cases}$$

For example, stock analysts often talk about moving averages of a stock price, such as, say, a 50 day moving average, to even out fluctuations.

[5.5]We have $d\tau = -dt$, but when we perform the substitution, we also have to interchange the limits of the integral, canceling the negative sign.

[5.6]Sometimes it helps clear conceptual understanding to note that two-way infinite sequences are just functions on $\mathbb{Z}$.

5.7 Problems

Problem 5.1. Let $X \geq 0$ be a real-valued random variable such that $E(X)$ exists and $P(X \neq 0) > 0$. Show that $E(X) \neq 0$.

Problem 5.2. Given two real-valued random variables X and Y, show that

$$\big(E(XY)\big)^2 \leq E(X^2) \, E(Y^2).$$

(This inequality is Schwarz's inequality for random variables).

Problem 5.3. Given two real-valued random variables X and Y with nonzero variances, show that

$$-1 \leq \mathrm{Corr}(X, Y) \leq 1.$$

The assumption that the variances of X and Y differ from zero is necessary in order that their correlation be defined.

6 Estimating time series parameters

Given a stationary time series $\{Y_t\}$, assume made observations y_k at times k with $1 \leq k \leq n$. It is natural to estimate $E(Y_t)$, which, under the assumption of stationarity, is independent of t, as

$$E(Y_t) \approx \bar{y} = \frac{1}{n} \sum_{k=1}^{n} y_k.$$

Such a procedure is not justified without further assumptions. Namely, we only made a single observation at time t, and estimating Y_t by observations made at different times does not necessarily give the correct result.

6.1 Convergence of random variables

Let $\langle X_n : 1 \leq n < \infty \rangle$ be a sequence of random variables, and let X be a random variable. We say that X_n converges to X in the squared mean if

$$\lim_{n \to \infty} E\big(|X - X_n|^2\big) = 0.$$

The absolute value is not needed if X and X_n are real valued. There are many other ways for a sequence of random variables to converge; for example, we say that X_n converges to X in the mean if

$$\lim_{n \to \infty} E\big(|X - X_n|\big) = 0;$$

however, convergence in the squared mean is technically easier to handle.

6.2 Ergodicity

A stationary time series $\{Y_t\}$ is called mean *ergodic* when the above procedure is justified, that is, when

$$E(Y_t) = \lim_{n \to \infty} \frac{1}{n} \sum_{m=0}^{n-1} Y_{t-m}.$$

Here, usually convergence in the squared mean is considered, in which case the process is called autocovariance ergodic in the squared mean.[6.1]

A stationary time series $\{Y_t\}$ is called autocovariance ergodic when

$$\mathrm{Cov}(Y_t, Y_{t-k}) = \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \big(Y_{t-m} - E(Y_{t-m})\big)\big(Y_{t-m-k} - E(Y_{t-m-k})\big) \qquad \text{for all} \quad k \in \mathbb{Z}.$$

Here, usually convergence in the squared mean is considered, in which case the process is called mean ergodic or autocovariance ergodic in the squared mean.[6.2]

7   The innovations algorithm

Let $k \in \mathbb{Z}$, and for each integer $r \geq k$ let Y_n be a random variable; for the sake of simplicity, assume Y_n is real valued, but these ideas can easily be modified so as to apply to complex-valued random variables. Assume, further, that $E(Y_n^2) < \infty$ for each $n \geq k$. Let

$$(7.1) \qquad\qquad \hat{Y}_t = \sum_{n=k}^{t-1} \alpha_{nt} Y_n.$$

for some coefficients α_{nt}. We say that $\hat{Y}_t$ is the best linear estimate for Y_t in terms of $\langle Y_n : k \leq n < t \rangle$ if for all choices of the coefficients α_n for n with $k \leq n < t$, with

$$(7.2) \qquad\qquad \tilde{Y}_t = \sum_{n=k}^{t-1} \alpha_n Y_n,$$

the expectation

$$(7.3) \qquad\qquad \mathrm{E}\big((Y_t - \tilde{Y}_t)^2\big)$$

is minimal if $\alpha_n = \alpha_{nt}$. We have[7.1]

Lemma 7.1. *Assume $\hat{Y}_t$ is the best linear estimate for Y_t in terms of $\langle Y_n : k \leq n < t \rangle$. Then we have*

$$\mathrm{E}\big((Y_t - \hat{Y}_t)Y_m\big) = 0$$

for all m with $k \leq m < t$.

[6.1]The term ergodic was introduced by Ludwig Boltzmann. Boltzman deduced the distribution of the speeds of molecules in a gas in equilibrium by studying the behavior of a small part of the gas through time. To make such a deduction possible, he had to assume that the time series associated with the behavior of a small part of the gas reflects the behavior of the whole volume of gas.

[6.2]Since $\{Y_t\}$ is assumed to be stationary here, the t on the right-hand side in the last two equations can be replaced with an arbitrary t'. One does not do this in a practical calculation, however, since the time series may only be approximately stationary in practice, so out of prudence one would use the same t on both sides of this equation.

[7.1]The geometric content of the lemma is that the shortest distance to a line or plane from an outside point is found by dropping a perpendicular on it. The quantity $\mathrm{E}(XY)$ is an inner product on the space of real-valued random variables on a given probability space, and this inner product creates a linear geometry. See Subsection 17.1 for a discussion of inner product spaces.

Proof. Let $\tilde{Y}_t$ be given by equation (7.2). We have

$$\mathrm{E}\big((Y_t - \tilde{Y}_t)^2\big) = E\Big(Y_t^2 - \sum_{n=k}^{t-1} 2\alpha_n Y_t Y_n + \sum_{l=k}^{t-1}\sum_{n=k}^{t-1} \alpha_l Y_l \alpha_n Y_n\Big)$$

$$= E\big(Y_t^2\big) - \sum_{n=k}^{t-1} 2\alpha_n \,\mathrm{E}(Y_t Y_n) + \sum_{l=k}^{t-1}\sum_{n=k}^{t-1} \alpha_l \alpha_n \,\mathrm{E}(Y_l Y_n).$$

To find the minimum of this, we take partial derivatives $\partial/\partial\alpha_m$ $(k \le m < t)$:

$$\frac{\partial}{\partial\alpha_m}\big(\mathrm{E}(Y_t - Y_t)^2\big) = -2\,\mathrm{E}(Y_t Y_m) + \sum_{l=k}^{t-1}\sum_{n=k}^{t-1}(\delta_{ml}\alpha_n + \alpha_l\delta_{mn})\,\mathrm{E}(Y_l Y_n)$$

$$= -2\,\mathrm{E}(Y_t Y_m) + 2\sum_{n=k}^{t-1}\alpha_n\,\mathrm{E}(Y_m Y_n)$$

The minimum is assumed when the right-hand side is zero, i.e., exactly when α_n for n with $k \le n < t$ is such that

$$\mathrm{E}\big((Y_t - \tilde{Y}_t)Y_m\big) = 0$$

for all m with $k \le m < t$. This completes the proof.[7.2] $\qquad\square$

It follows from the above proof that for $\hat{Y}_t$ to be the best linear estimate the coefficients on the right-hand side of (7.1) must satisfy the equations

(7.4)
$$\sum_{n=k}^{t-1}\alpha_{nt}\,\mathrm{E}(Y_m Y_n) = \mathrm{E}(Y_t Y_m) \qquad (k \le m < t).$$

7.1 Expressing the time series in terms of innovations

With the notation introduced above, write

$$e_t = Y_t - \hat{Y}_t$$

for $t \ge k$; to simplify the notation, we will assume $k = 0$ from now on. We call e_t the innovation at time t. Observe that we have

(7.5)
$$e_t = Y_t - \sum_{n=0}^{t-1}\alpha_{nt}Y_n \qquad (t \ge 0)$$

according to equation (7.1) with $k = 0$. It is easy in principle to solve these equations for Y_t in terms of e_n for $0 \le n \le t$; that is, we have

(7.6)
$$Y_t = \sum_{l=0}^{t}\psi_{l,t}e_{t-l} = \sum_{l=0}^{t}\psi_{t-l,t}e_l \qquad (t \ge 0)$$

[7.2]It is clear that the expression in (7.3) is a positive semidefinite form in the variables α_n, so it must have a minimum. So the equations we found describe the place of minimum.

with appropriate coefficients. It is also easy to see that for $t \geq 0$ in these equations we have

$$\text{(7.7)} \qquad\qquad \psi_{0,t} = 1.$$

Further, note that Lemma 7.1 implies that

$$\text{(7.8)} \qquad\qquad \text{E}(e_t e_n) = 0 \qquad \text{if} \quad 0 \leq n < t.$$

The innovations algorithm expresses the coefficients here in terms of the expectations $\text{E}(Y_l Y_m)$. For the sake of simplicity, we assume that for each $e_t \neq 0$ with positive probability for all $t \geq 0$; this implies that $\text{E}(e_t^2) > 0$ – see Problem 5.1.[7.3]

Multiplying equation (7.6) by e_n ($0 \leq n \leq t$) and taking expectations, and using (7.8), we obtain

$$\text{(7.9)} \qquad\qquad \text{E}(Y_t e_n) = \psi_{t-n,t}\,\text{E}(e_n^2) \qquad (t \geq 0).$$

To evaluate the coefficients $\psi_{t-m,t}$, we need to evaluate $\text{E}(Y_t e_n)$ and $\text{E}(e_n^2)$. This is fairly simple to do. Multiplying equation (7.6) by itself and taking expectations and taking (7.7) and (7.8) into account, we obtain that

$$\text{(7.10)} \qquad\qquad \text{E}(Y_t^2) = \text{E}(e_t^2) + \sum_{l=0}^{t-1} \psi_{t-l,t}^2\,\text{E}(e_l^2).$$

Finally, replacing t by m ($0 \leq m \leq t$) in equation (7.6), multiplying by Y_t, and taking expectations, using equation (7.7) we obtain

$$\text{E}(Y_t Y_m) = \sum_{l=0}^{m} \psi_{m-l,m}\,\text{E}(Y_t e_l) = \sum_{l=0}^{m} \psi_{m-l,m}\psi_{t-l,t}\,\text{E}(e_l^2). \qquad (0 \leq m \leq t).$$

Taking $n = t - m$ and omitting the middle member, this gives

$$\text{(7.11)} \qquad\qquad \text{E}(Y_t Y_{t-n}) = \sum_{l=0}^{t-n} \psi_{t-n-l,t-n}\psi_{t-l,t}\,\text{E}(e_l^2). \qquad (0 \leq n \leq t).$$

Assuming the mixed moments $E(Y_t Y_n)$ are known for all $t, n \geq 0$, equations (7.7), (7.10), and (7.11) can be used to evaluate the coefficients $\psi_{l,t}$ for $0 \leq l \leq t$ recursively. Equation (7.11) (with some help from equation (7.7)) is used to calculate $\psi_{n,t}$, and equation (7.11) is used to calculate $\text{E}(e_t^2)$. To be more specific, assume that $\psi_{n',t'}$ have been calculated for all pairs (n', t') such that $0 \leq t' < t$ and $0 \leq n' \leq t'$ or $t' = t$ and $n < n' \leq t$; also assume that $\text{E}(e_{t'}^2)$ has been calculated for all t' with $0 \leq t' < t$. Then we can calculate $\psi_{n,t}$ from the values calculated earlier, and in case $n = 0$ we can go on to calculate $\text{E}(e_t^2)$.

That is, we do the calculations in the following order: $\psi_{0,0}$, $\text{E}(e_0^2)$, $\psi_{1,1}$, $\psi_{0,1}$, $\text{E}(e_1^2)$, $\psi_{2,2}$, $\psi_{1,2}$, $\psi_{0,2}$, $\text{E}(e_2^2)$, $\psi_{3,3}$, $\psi_{2,3}$, $\psi_{1,3}$, $\psi_{0,3}$, $\text{E}(e_3^2)$, $\psi_{4,4}$, $\ldots$. See Problem 7.1 for details.

There is a cautionary note about the above formulas for calculating the coefficients $\phi_{t,l}$. They should be taken only as a theoretical description as to how to calculate these coefficients, and the formulas should not

---

[7.3]The algorithm that follows is essentially an adaptation of the Gram–Schmidt orthogonalization, discussed in Subsubsection 17.2.1. The additional complication here is that we do not normalize e_t here, i.e., we do not make the norm 1 in this case; that is, at present we usually do not have $\text{E}(e_t^2) = 1$. As for the requirement that $\text{E}(e_t^2) > 0$, it helps us to write the formulas in a simple way, but it is not essential. Gram–Schmidt orthogonalization has no difficulty with coping with occasional zero vectors – it simple skips them; see Subsection 17.2.1.

be used as a basis for writing numerical algorithms to calculate these coefficients. The problem is that the Gram–Schmidt orthogonalization, on which the above approach to calculate the coefficients $\phi_{t,l}$ is based, is numerically unstable. That is, small numerical errors committed initially (by rounding infinite decimals to a value representable on a computer) cause large errors later in the calculation. There are numerical methods avoiding these problem, and so there is no real impediment to calculate the coefficients $\phi_{t,l}$ accurately. This should be taken as a general comment for programming theoretical algorithms on computers: numerical analysis is a separate art, and theoretically correct algorithms may have to be modified when writing a computer implementation.

The recursive equations describing the innovations algorithm are discussed in [5, Proposition 5.2.2 on p. 165].

7.2 Zero means

In the discussion above, we did not assume that $\mathrm{E}(Y_t) = 0$, since there was no mathematical need to do so. When discussing time series, if it is possible to estimate the means of Y_t, it is natural to replace Y_t with $Y_t - \mathrm{E}(Y_t)$ as the first step in analyzing the time series. We will now make the assumption that

$$\mathrm{E}(Y_t) = 0 \qquad \text{for all} \quad t \in \mathbb{Z}.$$

Then an immediate consequence of equation (7.5) is that

$$\mathrm{E}(e_t) = 0 \qquad \text{for all} \quad t \in \mathbb{Z}.$$

In this case, one usually calls the innovation e_t at time as the error (committed by the mechanism producing the time series) at time t. One often also assumes that the variables e_t are independent normal variables for all t. Often there may be no rational reason to make this assumption other than the resulting ease of mathematical handling of the problem.

7.3 The partial autocorrelation function

Let $\{Y_t\}$ be a time series and let $X_t = Y_t - \mathrm{E}(Y_t)$. For a given t and $k \geq 0$, let $\hat{X}_{t+k}$ be the best linear estimate of X_{t+k} in terms of X_i with $t < i < t + k$ and let $\hat{X}_t$ be the best linear estimate of X_t in terms of the same X_i with $t < i < t + k$.

The definition of $\hat{X}_{t+k}$ is easy to understand in view of Lemma 7.1, and the innovations algorithm described in Subsection 7.1, and one can think of $\hat{X}_{t+k}$ as the value of X_{t+k} predicted in terms of X_i with $t < i < t + k$. The definition of $\hat{X}_t$ is somewhat less natural, since since X_t is known before one finds out the values of X_i with $t < i < t + k$. Nevertheless, the mathematics for this *postdiction*, i.e., "backward prediction," is the same, one merely needs to replace t by $N - t$ in the equations (for an arbitrarily chosen integer N – which can be 0 if one does not mind the fact that $-t$ may be a negative integer).[7.4]

The partial autocorrelation function of the time series $\{Y_t\}$ is defined as

$$\alpha_{t,t+k} = \mathrm{Corr}(X_t - \hat{X}_t, X_{t+k} - \hat{X}_{t+k}).$$

If Y_t is a stationary time series, $\alpha_{t,t+k}$ depends only on k, and not on t, and one may write $\alpha(k)$ instead of $\alpha_{t,t+k}$. Intuitively, $\alpha_{t,t+k}$ indicates the degree of relatedness between Y_t and Y_{t+k} with the intervening values of Y_i with $t \leq i \leq t + k$ removed.

[7.4]Replacing t by $-t$ (or by $N - t$) is called *time reversal*, discussed also below in Section 10.

7.4 Moving average models

Given a stationary process $\{Y_t\}$ we write

$$\gamma(n) = \mathrm{Cov}(Y_t, Y_{t-n});$$

the definition of stationarity given in Subsection 5.1 implies that the right-hand side here does not depend on t. It is also easy to see that $\gamma(n) = \gamma(-n)$. One also usually writes that

$$\rho(n) = \mathrm{Corr}(Y_t, Y_{t-n});$$

Note that $\rho(n)$ is defined unless $\gamma(0) = 0$, and the case $\gamma(0) = 0$ is of no interest.[7.5]

For $t \in \mathbb{Z}$ let e_t be uncorrelated random variables with zero means.[7.6] Let q be a positive integer. A moving average process $\{Y_t\}$ is a process of order q is a process satisfying the equations

$$(7.12) \qquad Y_t = e_t - \sum_{n=1}^{q} \theta_n e_{t-n}$$

with some coefficients θ_n for n with $1 \leq n \leq q$. Here e_t is called the error in the process at time t. If $\{Y_t\}$ is a stationary process such that $\mathrm{E}(Y_t) = 0$, then the innovation algorithm can be used to determine the coefficients θ_n. Writing

$$\theta(x) = 1 - \sum_{n=1}^{q} \theta_n x^n,$$

we can write

$$Y_t = \theta(B)e_t.$$

Corollary 4.2 requires that for all zeros λ of $\theta(x)$ we have $|\lambda| \geq 1$. Indeed, the random nature of the errors e_t will ensure that all basic of the solutions of the homogeneous equation $\theta(B)e_t$ will be represented in the solutions of the inhomogeneous equation $\theta(B)e_t = Y_t$ (considering this to be an equation of e_t for q initial values of Y_t, where q is the degree of $\theta(x)$. The solution corresponding to $|\lambda| < 1$ would imply that $\lim_{t \to \infty} e_t = \pm\infty$. This would also imply that $\lim_{t \to \infty} Y_t = \pm\infty$, and this would contradict that stationarity of Y_t.

7.5 Problem

Problem 7.1. Explain how equations (7.7), (7.9), (7.10), and (7.11) can be used to evaluate the moments $\mathrm{E}(e_t^2)$ and the coefficients $\psi_{t,l}$ for $0 \leq l \leq t$. assuming that the mixed moments $E(Y_t Y_n)$ are known for all $t, n \geq 0$.

8 Autoregressive processes and the Yule–Walker equations

Assume $\mathrm{E}(Y_t) = 0$ for all t. The process $\{Y_t\}$ is said to be autoregressive of order p if the following conditions are satisfied. The

$$(8.1) \qquad\qquad Y_t = \sum_{k=1}^{p} \phi_k Y_{t-k} + e_t \qquad (t \in \mathbb{Z})$$

[7.5]$\gamma(0) = 0$, i.e., $\mathrm{Var}(Y_t) = 0$ means that Y_t is almost surely constant (see Problem 5.1. Since $\mathrm{E}(Y_t)$ does not depend on t for stationary series, this means that the whole series $\{Y_t\}$ almost surely assumes the same value.

[7.6]One often assumes that the variables e_t are independent normal variables; their variances do not have to be the same.

holds, where e_t is the error at time t; it is assumed that the random variables e_t are uncorrelated, and $\mathrm{E}(e_t) = 0$ and $\sigma(e_t) < \infty$ for all t; the variable e_t is unobservable. Further, we assume that for all $t' < t$, the variable $Y_{t'}$ is uncorrelated with e_t.

Writing

$$(8.2) \qquad \phi(x) = 1 - \sum_{k=1}^{p} \phi_k x^k$$

and using the backshift operator, we have

$$\phi(B)Y_t = e_t.$$

Assuming $\{Y_t\}$ is a stationary process, given an arbitrary integer (positive, negative, or zero), the covariance $\gamma_{t,t-k} = \mathrm{Cov}(Y_t, Y_{t-k})$ does not depend on k, so we can write $\gamma(k) \stackrel{def}{=} \gamma_{t,t+k}$. This assumption allows us to derive a system of equations for the coefficients ϕ_k in equation (8.1). Hence, for $k > 0$ we have

$$\gamma(k) = \mathrm{Cov}(Y_t, Y_{t-k}) = \mathrm{Cov}\Big(\sum_{i=1}^{p} \phi_i Y_{t-i} + e_t, Y_{t-k}\Big)$$

$$= \sum_{k=1}^{p} \phi_i \, \mathrm{Cov}(Y_{t-i}, Y_{t-k}) + \mathrm{Cov}(e_t, Y_{t-k}) = \sum_{i=1}^{p} \phi_i \gamma(k-i);$$

the last equation holds since $\mathrm{Cov}(e_t, Y_{t-k}) = 0$ for $k > 0$ (the assumption $k > 0$ is essential here, since the error e_t at time t certainly influences the value of Y_t). That is

$$(8.3) \qquad \gamma(k) = \sum_{i=1}^{p} \phi_i \gamma(k-i)$$

for any integer $k > 0$. Noting that with $\rho_{t,t-k} \stackrel{def}{=} \mathrm{Corr}(Y_t, Y_{t-k})$, $\rho(k) = \rho_{t,t-k}$ does not depend on t, and $\rho(k) = \gamma(k)/\gamma(0)$, and we can divide the above equation with $\gamma(0)$. Taking these equations only for k with $1 \leq k \leq p$, we obtain.

$$(8.4) \qquad \rho(k) = \sum_{i=1}^{p} \phi_i \rho(k-i) \qquad (1 \leq k \leq p).$$

These equations are the equivalent equations (8.3) are called the Yule–Walker equations.

While the derivation shows that these equations should also be satisfied for $k > p$, but then we may have more equations than unknowns, and the equations may be contradictory. For the optimal choice of p, taking a $p' > p$, in the analogous equations

$$\rho(k) = \sum_{i=1}^{p'} \phi_i' \rho(k-i) \qquad (1 \leq k \leq p'),$$

ϕ_i' should not be significantly different from ϕ_i for k with $1 \leq i \leq p$ and ϕ_i' should not be significantly different from 0 for $i > p$.

8.1 Best linear prediction for stationary processes

Given a stationary time series $\{Y_t\}$, we want to predict Y_t in terms of the previous p observations $\{y_{t-n}\}_{1 \leq n \leq p}$, and write

$$\hat{Y}_{t-1}(1) = \sum_{n=1}^{p} \hat{\phi}_n y_{t-n};$$

the symbol on the left-hand side denotes the one-step ahead prediction of Y_t made at time $t - 1$.[8.1] That is, writing

$$Y_{t-1}(1) = \sum_{n=1}^{p} \hat{\phi}_n Y_{t-n},$$

$\hat{Y}_{t-1}(t)$ is the value obtained for the random variable $Y_{t-1}(1)$ by substituting the observed values of $\{Y_{t-n}\}_{1 \leq n \leq p}$ in this equation. The prediction error is the random variable $Y_t - \hat{Y}_{t-1}(1)$. Because of the assumption that $\mathrm{E}(Y_t) = 0$ for all t made above, we can see that the mean square prediction error is

$$(8.5) \qquad \mathrm{Var}\big(Y_t - Y_{t-1}(1)\big) = \mathrm{E}\left(\left(Y_t - \sum_{n=1}^{p} \hat{\phi}_n Y_{t-n} \right)^2 \right).$$

This will be minimum when equations (7.4) are satisfied with $\hat{\phi}_{t-n}$ replacing α_{nt} and $t - p$ replacing k. Noting that $\{Y_t\}$ is stationary with mean 0, we have $\mathrm{E}(Y_{t-i}Y_{t-n}) = \mathrm{Cov}\,\mathrm{E}(Y_{t-i}Y_{t-n}) = \gamma(n - i)$, and similarly, $\mathrm{E}(Y_t Y_{t-i}) = \gamma(i)$, this gives the equations

$$(8.6) \qquad \sum_{n=1}^{p} \hat{\phi}_n \gamma(i - n) = \gamma(i) \qquad (1 \leq i \leq p).$$

These equations are identical to the Yule–Walker equations (8.3) given above.

8.2 Solvability of the Yule–Walker equations

We have the following

Theorem 8.1. *Let $\{Y_t\}$ be a stationary process such that $\gamma(0) \neq 0$ and $\lim_{n \to \infty} \gamma(n) = 0$. Then, for every $p \geq 1$, the covariance matrix $\Gamma_p \overset{def}{=} \big(\gamma(k - i)\big)_{1 \leq i,k \leq p}$ is nonsingular.*

The covariance matrix Γ_p is the matrix of the form of the Yule–Walker equations given in (8.3)); the nonsingularity of this matrix means that those equations or equations (8.3) have a unique solution. For the proof, we need the following

Lemma 8.1. *Let $m \geq 1$, and X_i for i with $1 \leq i \leq m$ be random variables such that the covariance matrix $C = \big(\mathrm{Cov}(X_i, X_j)\big)_{1 \leq i,j \leq m}$ is singular. Then there is an r with $0 \leq r < m$ and there are numbers a_1, a_2, ..., a_r such that $X_{r+1} = \sum_{i=1}^{r} a_i X_i$ with probability 1.*

Further, if X_i' for i with $1 \leq i \leq m$ is another collection of random variables with the same covariance matrix C, then we have $X_{r+1}' = \sum_{i=1}^{r} a_i X_i'$ with the same r and the same coefficients a_i for $1 \leq i \leq r$.

[8.1]$\hat{Y}_{t-1}(1)$ is more or less standard notation for the one-step ahead prediction made at time $t - 1$. In equation (7.1) used a different notation, since the present notation would have been too cumbersome in those considerations.

The lemma can be found in [5, Proposition 5.1.1 on p. 160].

Proof. As C is singular, there is an r with $0 \leq r < m$ such that the $(r+1)$st row of C is a linear combination of its earlier columns; that is

$$\text{Cov}(X_{r+1}, X_j) = \sum_{i=1}^{r} a_i \, \text{Cov}(X_i, X_j) \qquad \text{for all } j \text{ with } 1 \leq j \leq m$$

Hence

$$0 = \text{Cov}(X_{r+1}, X_j) - \sum_{i=1}^{r} a_i \, \text{Cov}(X_i, X_j) = \text{Cov}\Big(X_{r+1} - \sum_{i=1}^{r} a_i X_i, \ X_j\Big) \qquad (1 \leq j \leq m).$$

Any linear combination of the right-hand sides of these equations also gives 0. Thus

$$\text{Cov}\Big(X_{r+1} - \sum_{i=1}^{r} a_i X_i, \ X_{r+1} - \sum_{i=1}^{r} a_i X_i\Big) = 0.$$

Therefore, the existence of an r as claimed follows. As for the last sentence of the lemma, it follows since the covariance matrix C by itself allowed us to find r and the coefficients a_i. □

For the proof, we need some background about matrices. An orthogonal $r \times r$ matrix is such that $Q^T Q = I$, where I is the $r \times r$ identity matrix. This means that Q^T is the left inverse of Q. If a square matrix has a left inverse, then it also has the right inverse, and it is the same as the left inverse. Hence we also have $QQ^T = I$ for an orthogonal matrix. The l^2 norm of an r-dimensional column vector $\mathbf{x} = (x_1, x_2, \ldots, x_r)^T$ is defined as

$$\|\mathbf{x}\| = \sqrt{\sum_{i=1}^{r} x_i^2}.$$

This norm is sometimes also denoted as $\|\mathbf{x}\|_2$ to indicate that we are talking about l^2 norms, but we will refrain from this, since the only vector norm we will use is the l^2 norm. An $r \times r$ matrix A has a norm induced by the given vector norm:

$$\|A\| = \max\{\|A\mathbf{x}\| : \|\mathbf{x}\| = 1\}.$$

We have $\|\mathbf{x}\|^2 = \mathbf{x}^T x$. If Q is an orthogonal matrix, we have

(8.7) $\qquad \|Q\mathbf{x}\|^2 = (Q\mathbf{x})^T (Q\mathbf{x}) = \mathbf{x}^T Q^T Q\mathbf{x} = \mathbf{x}^T (Q^T Q)\mathbf{x} = \mathbf{x}^T I\mathbf{x} = \mathbf{x}^T \mathbf{x} = \|x\|^2.$

This shows that an orthogonal matrix preserves vector norms; for this reason, it is also an isometry (i.e., it preserves distances, i.e., the metric). For more about orthogonal matrices, see see [23, §38, p. 175-176].[8.2]

Proof of Theorem 8.1. Assume there is a $p \geq 1$ such that Γ_p is singular, and let p be the smallest such integer. Then $p > 1$ since Γ_1 is the 1×1 matrix with $\gamma(0)$ as its only entry, and $\gamma(0) \neq 0$ by

---

[8.2]The quoted pages are interesting for statistics also in other respects, since §38 discusses the numerical handling of the least squares approximation.

our assumption. Let $k \geq p$ be an integer. Applying the lemma $m = p$ for the random variables Y_{k-p+i} with $1 \leq i \leq k$, we find an r with $1 \leq r < p$ and numbers a_i for $1 \leq i \leq r$ such that

$$Y_{k-p+r+1} = \sum_{i=1}^{r} a_i Y_{k-p+r}.$$

Note that r does not depend on k in view of the last sentence of the lemma. The only important point here is that $Y_{k-p+r+1}$ can be expressed as a linear combination of Y_j for j with $1 \leq j \leq k - p + r$. Applying this result for each term on the right-hand side, we obtain that $Y_{k-p+r+1}$ is expressible as a linear combination of Y_j for $1 \leq j \leq k - 1 - p + r$ as long as $k - 1 \geq p$. Repeating this argument $k - p$ times, we obtain that $Y_{k-p+r+1}$ is expressible as as a linear combination of Y_j with $1 \leq j \leq r$.

Taking $n = k - p + r + 1$, we obtain

(8.8)
$$Y_n = \sum_{i=1}^{r-1} a_i^{(n)} Y_j$$

for every $n > r$, where the coefficients $a_i^{(n)}$ may depend on n. Multiplying this equation by Y_n and taking expectations, we obtain that

$$\gamma(0) = \sum_{i=1}^{r-1} a_i^{(n)} \gamma(n - j).$$

Making $n \to \infty$, we have $\gamma(n - j) \to 0$ by our assumptions, and so, writing $\mathbf{a}^{(n)} = (a_1^{(n)}, (a_2^{(n)}, \ldots, a_r^{(n)})^T$, we must have

(8.9)
$$\lim_{n \to \infty} \|a^{(n)}\| = +\infty$$

for the last equation to hold.

Writing $\mathbf{Y}_k$ for the column vector $(Y_{k+1}, Y_{k+2}, \ldots, Y_{k+r})^T$ equation (8.8) can be written in matrix form as

$$Y_n = \mathbf{Y}_1^T \mathbf{a}^{(n)}.$$

Multiplying by the column vector $\mathbf{Y}_1^T$ on the left and taking expectations, we obtain that

(8.10)
$$(\gamma_{n-1}, \gamma_{n-2}, \ldots, \gamma_{n-r})^T = \Gamma_r \mathbf{a}^{(n)}.$$

The matrix Γ_r is a symmetric positive semi-definite, and since it is nonsingular by the minimality assumption of p (p was assumed to be the smallest integer such that Γ_p is singular, and $r < p$), it follows that Γ_r is positive definite. All eigenvalues of a positive definite symmetric matrix are positive real numbers. By the Principal Axis Theorem of linear algebra, there is an orthogonal $r \times r$ matrix Q such that

$$\Gamma_r = Q^T D Q,$$

where D is a diagonal matrix with the eigenvalues λ_1, λ_2, ..., λ_r of Γ_r being its diagonal entries; we may assume that the entries occur in increasing order; that is, $D = (\lambda_i \delta_{ij})_{1 \leq i,j \leq r}$ with $0 < \lambda_1 < \lambda_2 < \ldots < \lambda_r$; see e.g., [30, Theorem 7.4.4' on p. 333].

Hence, for the norm of the right-hand side of (8.10) we have

$$\|\Gamma_r \mathbf{a}^{(n)}\| = \|Q^T D Q \mathbf{a}^{(n)}\| = \|D Q \mathbf{a}^{(n)}\| \geq \lambda_1 \|Q \mathbf{a}^{(n)}\| = \|\mathbf{a}^{(n)}\|.$$

The second and third equalities here hold since Q^T and Q are both orthogonal matrices, and so they preserve norms (cf. (AR: preserve norms), and the inequality holds since for a column vector $\mathbf{x} = (x_1, x_2, \ldots, x_r)^T$ we have $D\mathbf{x} = (\lambda_1 x_1, \lambda_2 x_2, \ldots, \lambda_r x_r)^T$. Since the left-hand side of (8.10) tends to zero as $n \to \infty$, we must have $\lim_{n \to \infty} \|\mathbf{a}^{(n)}\| = 0$. This contradicts (8.9), completing the proof. $\qquad\square$

8.3 Solving the Yule–Walker equations

The matrix Γ_p has a special form, and so in solving equations (8.3) there are methods faster than Gaussian elimination. The matrix Γ_p is a *Toeplitz matrix*, that is, a square matrix in which all elements is a diagonal parallel to the main diagonal are the same. That, is a square matrix (t_{ij}) is a Toeplitz matrix if $t_{ij} = t_{i'j'}$ whenever $i - j = i' - j'$. Systems of linear equations whose matrix is a Toeplitz matrix can be solved by variants of the Levinson algorithm – see [41]

8.4 Location of zeros of the autoregressive model polynomials

Assume $\{Y_t\}$ is a stationary process of order p as described at the beginning of Section 8. Then, using the backshift operator B for the covariances to mean $B\gamma(k) = \gamma(k-1)$, equation (8.3) can be described with the aid of the polynomial $\phi(x)$ given in (8.2) as

$$\phi(B)\gamma(k) = 0;$$

this equation is true for all $k > 0$, even though in stating the Yule-Walker equations, we restricted k to the range $1 \leq k \leq p$. Considering this as a homogeneous recurrence equation for $\gamma(k)$, its solutions are linear combinations of the basic solutions given in Corollary 4.2. In an example occurring in practice, the solution is represented by a linear combination in which the coefficient of any basic solution is nonzero, since there need to be very special initial conditions to ensure that such a coefficient is zero. If $\phi(x)$ has a zero λ with $|\lambda| \leq 1$, then this ensures that $\lim_{k \to \infty} \gamma(k) \neq 0$. This is an undesirable behavior in a stationary time series, therefore, in autoregressive models one usually requires that $|\lambda| > 1$ for all zeros λ of $\phi(x)$.

9 Mixed autoregressive moving average processes

9.1 ARMA models

Let $\{Y_t\}$ be a stationary time series such that $E(Y_t) = 0$ for all $t \in \mathbb{Z}$. The process $\{Y_t\}$ is called a mixed autoregressive moving average process of order (p, q) if

$$(9.1) \qquad Y_t = \sum_{k=1}^{p} \phi_k Y_{t-k} + e_t - \sum_{k=1}^{q} \theta_k e_{t-k},$$

where e_t is the error at time t; it is assumed that the random variables e_t are uncorrelated, and $E(e_t) = 0$ and $\sigma(e_t) < \infty$ for all t. The error e_t is also uncorrelated with $Y_{t'}$ with $t' < t$. Note that the variable e_t is unobservable. Such a process is also called an ARMA(p, q) process. Writing

$$(9.2) \qquad \phi(x) = 1 - \sum_{k=1}^{p} \phi_k x^k \quad \text{and} \quad \theta(x) = 1 - \sum_{k=1}^{q} \theta_k x^k,$$

we have

$$(9.3) \qquad \phi(B)Y_t = \theta(B)e_t.$$

This model is reducible if the greatest common divisor $\chi(x)$ of $\phi(x)$ and $\theta(x)$ is not constant (i.e., if it has degree greater than or equal to 1), since in that case we can divide both sides by $\chi(x)$. In any case, the model can also be written as

$$(9.4) \qquad Y_t = \frac{\theta(B)}{\phi(B)}e_t,$$

or

$$\frac{\phi(B)}{\theta(B)}Y_t = e_t.$$

Replacing the fractions by their Taylor series at 0, these represent a pure moving average process and a pure autoregressive process of infinite order, respectively. For the convergence of these series certain assumptions are needed. We do not quite need to assume that the process $\{Y_t\}$ is stationary. Assuming that the expectations of the squares of Y_t and e_t are bounded, i.e., that there is a number M such that $\mathrm{E}(Y_t^2) < M$ and $\mathrm{E}(e_t^2) < M$,[9.1] and the power series

$$f(x) = \sum_{k=0}^{q} a_k x^k,$$

has radius of convergence greater than one, the series $f(B)Y_t$ and $f(B)e_t$ converge in the mean (i.e., in expectation); for example, for Y_t this means that there is a random variable $\bar{Y}_t$ such that

$$\lim_{n\to\infty} \mathrm{E}\left(\left|\bar{Y}_t - \sum_{k=0}^{n} a_k Y_{t-k}\right|\right) = 0.$$

This follows from known results of integration theory, somewhat beyond the scope of this course. Instead of convergence in the mean, one often prefers convergence in the square mean for technical reasons:

$$\lim_{n\to\infty} \mathrm{E}\left(\left|\bar{Y}_t - \sum_{k=0}^{n} a_k Y_{t-k}\right|^2\right) = 0;$$

the absolute value is unnecessary for real-valued random variables.

If one requires that the polynomials $\phi(x)$ and $\theta(x)$ have no zeros in the closed unit disk $\{z \in \mathbb{C} : |z| \leq 1\}$ of the complex place $\mathbb{C}$, the radius of convergence of the Taylor series of both $\phi(x)/\theta(x)$ and $\theta(x)/\phi(x)$ will be greater than 1; see Section 3.

9.2   Coefficients in the pure MA representation

Writing $\psi(x) = \theta(x)/\phi(x)$, equation (9.4) can be written as

$$(9.5) \qquad Y_t = \psi(B)e_t,$$

where $\psi(x)$ can be represented as an infinite series

$$(9.6) \qquad \psi(x) = \sum_{n=0}^{\infty} \psi_n x^n.$$

[9.1]It follows from simple inequalities involving expectations that it is enough for this that the both the mean and the variance of Y_t and e_t are bounded.

Assuming the zeros of $\phi(x)$ are outside the unit circle, this series has radius of convergence > 1. The coefficient ψ_n can easily be determined from the coefficients in equations (9.2). Indeed, writing $\phi_0 = \theta_0 = -1$ and $\psi_n = 0$ for $n < 0$, the equation $\psi(x)\phi(x) = \theta(x)$ can be written as

$$\sum_{n=0}^{\infty} \Big(\sum_{k=0}^{p} \psi_{n-k}\phi_k\Big) x^n = \sum_{n=0}^{q} \theta_n x^n.$$

Equating the coefficients of x^n gives equations for the coefficients ψ_n. That is, noting that $\phi_0 = -1$, we have

(9.7) $$\psi_n = -\theta_n + \sum_{k=1}^{p} \psi_{n-k}\phi_k \qquad (n \geq 0),$$

where we take $\theta_n = 0$ for $n > q$.[9.2]

9.3 Calculating the MA coefficients in ARMA models

Given a stationary time series $\{Y_t\}$ with zero means and positive integers p, q, we would like to build an ARMA(p, q) model described in equation (9.1). Assume the coefficients ϕ_k for k with $1 \leq k \leq p$ have already been determined. The question is how to determine the coefficients θ_i for k with $1 \leq k \leq q$. To do this, we first need to build a pure MA model, using the innovation algorithm described in Subsections 7.1 to calculate the coefficients in a pure MA model (9.5), with the infinite series in equation (9.6) truncated to a finite sum:

$$Y_t = \sum_{l=0}^{m} \psi_n e_{t-n}$$

for some integer m, where $\psi_0 = 1$; cf. equation (7.6); at present, the coefficients ψ_l do not depend on t, since the time series $\{Y_t\}$ is stationary. Choosing larger values of m will give more accurate results; in any case, we need to make sure that $m \geq \max(p, q)$. Then, using equations (9.7), the coefficients θ_n for n with $0 \leq n \leq q$ can be determined.[9.3]

9.4 The primacy of autoregressive models

Moving average models are a kind of mathematical artifact, and they do not reflect natural forces producing the time series. An error committed at an earlier time does not directly govern the present behavior of the time series $\{Y_t\}$. Any effect on the present value of Y_t is brought about by the earlier errors is mediated through the values of $Y_{t'}$ for $t' < t$. This means that autoregresssive models give a natural description of the time series via an equation of the form

(9.8) $$\phi_{\infty}(B)Y_t = \epsilon_t,$$

where the ϵ_t is the error committed at the present time, while the subscript of ϕ indicates that

$$\phi_{\infty}(x) = 1 - \sum_{k=1}^{\infty} \phi_k x^k$$

[9.2]These equations just express in equations what happens when we perform the usual long division $\theta(x)/\phi(x)$ with the modification that the powers of x are arranged in reverse order, i.e., in the order 1, x, x^2, ..., and the process goes on indefinitely.

[9.3]We have $\theta_0 = -1$ according to these equations, since $\psi_0 = 0$; θ_0 does not occur directly in equation (9.1).

is usually an infinite series.[9.4] If the time series $\{Y_t\}$ is stationary, it is natural to think about B as an operator of norm 1,[9.5] and so, for the convergence of the series on the left of (9.8), one wants to make sure that the radius of convergence of the series $\phi_\infty(x)$ is greater than 1.

In numerical calculations, one truncates $\phi_\infty(x)$ to a polynomial

$$\phi_m(x) = 1 - \sum_{k=1}^{m} \phi_k x^k,$$

and considers the truncated model

(9.9) $$\phi_m(B)Y_t = e_t,$$

Truncation causes the errors e_t on the right-hand side to be different from ϵ_t, but if we choose m to be large, e_t will be a good approximation to ϵ_t. Assuming Y_t has zero means, we have

$$\mathrm{Var}(e_t) = \mathrm{E}\Big(\big(\phi_m(B)Y_t\big)^2\Big),$$

where, for stationary $\{Y_t\}$, the right-hand side does not depend on t. So the variance of e_t is independent of t. Dividing equation (9.9) by $\phi_m(B)$, we obtain

(9.10) $$Y_t = \frac{1}{\phi_m(B)} e_t,$$

Here $1/\phi_m(x)$ can be written as a power series. In view of Subsection 8.4, it seems reasonable to assume that $\phi_m(x)$ has no zeros λ with $|\lambda| \leq 1$, so the Taylor series for $1/\phi_m(x)$ has radius of convergence > 1. Hence the series for $1/\phi_m(B)$ is convergent when applied to e_t.[9.6]

When building an ARMA model, $1/\phi_m(x)$ in equation (9.10) is not calculated from an AR model; and an approximation to $1/\phi_m(x)$ is obtained via the innovation algorithm of Section 7 directly from the autocovariances $\gamma(k)$ of the stationary time series $\{Y_t\}$. In any case, $1/\phi_m(x)$ has no zeros anywhere, it being a reciprocal. Further, one takes an AR model

$$\phi(B)Y_t = e_t,$$

where $\phi(x)$ is a further truncation of $\phi_m(x)$. Then one replaces the model (9.10) with the equivalent model

$$\phi(B)Y_t = \frac{\phi(B)}{\phi_m(B)} e_t.$$

Finally, one takes a polynomial approximation $\theta(x)$ of the infinite series $\phi(x)/\phi_m(x)$, and the sought-after ARMA model will be

$$\phi(B)Y_t = \theta(B)e_t.$$

In fact, we can take $\theta(x)$ to be a truncation to a polynomial of the infinite power series $\phi(x)/\phi_m(x)$.

[9.4]We mentioned above in Subsection 5.2 that such equations are naturally satisfied by time series produced by simple state-space models described in Section 23.

[9.5]Norms are systematically discussed only in Subsection 17.1 below, since initially we want to avoid too much abstract discussion, but at present, by the norm of a random variable X we mean $\|X\| \stackrel{def}{=} \big(\mathrm{E}(|X|^2)\big)^{1/2}$, and B having norm 1 means that $\|B(Y_t)\| = \|Y_t\|$ in view of stationarity.

[9.6]As we pointed out above, $\mathrm{E}(e_t^2)$ does not depend on t, so $\|B(e_t)\| = \|e_t\|$.

9.4.1 Why the MA polynomial in an ARMA model is expected to have no zeros in the closed unit disk

As $\phi(x)$ is usually assumed to have no zeros λ with $|\lambda| \le 1$ in view of Subsection 8.4, neither will have $\phi(x)/\phi_m(x)$ have such zeros (also recall that $1/\phi_m(x)$ is represented by a power series with radius of convergence greater than 1). If $\theta(x)$ is a good enough approximation of $1/\phi_m(x)$, then $\theta(x)$ will have no zeros in the closed unit disk either.

9.5 Prediction with ARMA models

Assume we have observations y_t of Y_t for $0 \le t < n$, and that these observations satisfy equation (9.1), or equation (9.3), which is just a short form of the former equation, with the appropriate changes (such as replacing Y_t by y_t). We want to predict Y_n at time $n - 1$; we will denote the prediction with $\hat{Y}_n(1)$. To this end we write $\hat{e}_t$ for the estimated error at time t with $-\infty < t < n$. These error estimates are obtained by solving the equation

$$\theta(B)\hat{e}_t = \phi(B)y_t \qquad \left(p \le t < n\right),$$

with initial conditions $\hat{e}_t = 0$ for $-\infty < t < p$. Note that this is an inhomogeneous recurrence equation for $\hat{e}_t$, since the right-hand side is known. If we change the initial conditions to the actual values of the error e_t for $t < p$ then the change in the solution for e_t will be a solution of the homogeneous equation (i.e., the above equation with 0 right-hand side). If we require that all zeros of $\theta(x)$ are outside the unit circle, then every solution of the homogeneous equation will tend to 0 as $t \to \infty$, according to Corollary 4.2, so, assuming n is large, taking 0 as initial condition will ensure that the estimates $\hat{e}_t$ will be close to the actual value of e_t for $n - q \le t < n$. We put

$$\hat{Y}_{n-1}(1) = \sum_{k=1}^{p} \phi_k y_{n-k} - \sum_{k=1}^{q} \theta_k \hat{e}_{n-k}.$$

This is is just equation (9.1), with n replacing t, $\hat{Y}_{n-1}(1)$ replacing Y_t on the left-hand side, $\hat{Y}_{n-k}$ replacing Y_{t-k} for k with $1 \le k \le p$, $\hat{e}_{n-k}$ replacing e_{t-k} for k with $1 \le k \le q$, and 0 replacing e_t on the right-hand side.

9.6 The importance of ARMA models

As we saw in Subsection 9.1, and ARMA model can also be written as a pure autoregressive model, or a pure moving average model, each with possibly infinitely many coefficients. The importance of ARMA models lies in that they allow to model the time series with fewer parameters.

9.7 Integrated ARMA models

Let $\{Y_t\}$ be a time series, and assume that using the difference operator $\nabla = I - B$ d times, where $d \ge 0$ is an integer, we arrive at the time series $\{(I - B)^d Y_t\}$ that is a stationary time series with 0 means. Then we can model the latter time series by an ARMA(p, q) model, that is we can write

(9.11) $$\phi(B)(I - B)^d Y_t = \theta(B)e_t$$

according to equation (9.3), where $\phi(x)$ and $\theta(x)$ are are as in equation (9.2). If p and q are the degrees of $\phi(x)$ and $\theta(x)$, respectively, such an equation is called an integrated autoregressive moving average model of order (p, d, q), or, shortly, and ARIMA(p, d, q) model.

10 Time reversal

When studying a time series

$$\{Y_t\} = \langle Y_t : t \in \mathbb{Z} \rangle,$$

is it sometime useful to also look at the time series obtained by *time reversal*, i.e., the time series in which t is replaced by $-t$:

$$\{Y_{-t}\} = \langle Y_t : t \in \mathbb{Z} \rangle.$$

There may be various mathematical justification for studying the time reversed series. For example, if the time series $\{Y_t\}$ is stationary, the time reversed series $\{Y_{-t}\}$ is also stationary, with the same covariance coefficients. Hence, when constructing an autoregressive model for the time series $\{Y_t\}$, the same autoregressive model also works for the time reversed series $\{Y_{-t}\}$ in view of the Yule–Walker equations (equations (8.3) or (8.4)). Since ARMA models can naturally related to autoregressive models (cf. equation (9.4), this observation also extends to ARMA models.

Time reversal is also important in physics when studying time-reversal symmetric equations. Yet, in a sense one feels uneasy about time reversal, since one has never seen a broken coffee cup spontaneously reassemble its pieces into a whole coffee cup. Much ink has been spilled on physico-philosophical explanations why this does not happen in spite of the time-reversal symmetry of the equations of physics, but none of these explanations seem truly convincing.

Similarly, one may feel uneasy about time reversal in time series, since it is natural to attribute a random component to future events; it is much less natural to attribute randomness to past events. In any case, if the mathematical theory works, why not make use of it. The doubt however persists whether a given time series, especially one obtained by differencing, can really be described as stationary.

10.1 Estimating the residuals of an ARMA model

Assuming that the time series $\{Y_t\}$ has the ARMA model

(10.1) $$\phi(B)Y_t = \theta(B)e_t$$

(cf. (9.3), the same ARMA model

$$\phi(B)Y_{-t} = \theta(B)e_{-t}$$

for the reversed time series can also be written as

(10.2) $$\phi(E)Y_t = \theta(E)e_t,$$

where E is the forward shift operator (see Subsection 4.1). Assume we have observations y_t of Y_t for $0 \le t \le n$, and that these observations satisfy equation (10.1) for an ARMA(p, q) model. We have described in Subsection 9.5, given initial values of the errors e_t, usually called *residuals* in the time series literature, for t with $0 \le t < p$, we can calculate the residuals for t with $p \le t \le n$. Similarly, if the values of e_t for t with $n - p < t \le n$ are given, using equation (10.2) we can calculate e_t for t with $0 \le t \le n - p$. As we also pointed out in Subsection 9.5, if the zeros of $\theta(x)$ are outside the unit circle (as required for ARMA models), then errors committed in the initial values for e_t die out as t increases.

This motivates the following procedure to determine the residuals (see [4, Section 7.1.5, pp. 233–235]). For a start, take $\hat{e}_{t,1} = 0$ for t with $0 \le t < p$ and, using equation (10.1), calculate $\hat{e}_{t,1}$ for t with $p \le t \le n$. Next use the initial values $\hat{e}_{t,2} = \hat{e}_{t,1}$ for t with $n - p < t \le n$ with equation equation (10.2) to determine $\hat{e}_{t,2}$ for t with $0 \le t \le n - p$. To go forward, use the initial values

$\hat{e}_{t,3} = \hat{e}_{t,2}$ for t with $0 \leq t < p$ with equation equation (10.1), to calculate $\hat{e}_{t,3}$ for t with $p \leq t \leq n$. Assuming that $\theta(x)$ has all its zeros outside the unit cirle, this procedure converges, and we can take

$$\hat{e}_t = \lim_{k \to \infty} \hat{e}_{t,k} \qquad (0 \leq t \leq n)$$

for the values of the residuals of the observed time series $\{y_t\}$.

10.2 Conditional and unconditional sum of squares

Given the above observed time series $\langle y_t : 0 \leq t \leq n \rangle$, the *conditional sum of squares of the residuals* is the sum

$$\sum_{t=0}^{n} \hat{e}_{t,1}^2.$$

That is, this sum is conditional on the assumption that the initial values of the residuals are taken to be 0 in the above calculation.[10.1] The unconditional sum of squares is the sum

(10.3)
$$\sum_{t=0}^{n} \hat{e}_t^2,$$

where $\hat{e}_t$ can be calculated as described above. The coefficients of the model are usually described by requiring that the conditional sum of squares or else the unconditional sum of squares be the least possible. The method relying on the conditional sum of squares is numerically more stable, but, especially since a short time series there is not enough time for the errors in the initial conditions to die out,[10.2] a more accurate model may be constructed by using the unconditional sum of squares. It also seems that using the unconditional sum of squares method strongly relies on the correctness of the assumption that the time series is stationary, while this is not the case for the conditional sum of squares method.

10.3 The likelihood function of an ARMA model

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a vector of random variables, and assume we are considering a model in which the joint density function of $\mathbf{X}$ is described as as a function $f(\mathbf{P}, \mathbf{x})$, where $\mathbf{P} = (P_1, P_2, \ldots, P_m)$ is a vector of parameters. The goal is to find the parameter vector best describing the random variable vector $\mathbf{X}$. Assume we have a single observation x_i for the random variable X_i for $1 \leq i \leq n$. When one considers these observations $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ as given, $f(\mathbf{P}, \mathbf{x})$ is called the likelihood function of the parameter vector $\mathbf{P}$. The *Maximum Likelihood Estimate* takes the place of maximum $\mathbf{P} = \mathbf{P}^{(0)}$ of the function $f(\mathbf{P}, \mathbf{x})$ as the estimate of the parameters $\mathbf{P}$ of the model.[10.3]

[10.1]The initial condition 0 seems reasonable in that nothing is known about the residuals except that their expectation is 0.

[10.2]This is especially so if in the obtained model $\theta(x)$ has a zero that is close to the unit circle (while being outside the unit circle).

[10.3]The nature of most statistical models is such that the function $f(\mathbf{P}, \mathbf{x})$ has a single place of maximum for fixed $\mathbf{x}$. One would be tempted to describe such a model as *unimodal*, but one needs to be somewhat cautious here, since the term "unimodal" is usually used for density functions having a single place of maximum, and we are talking about likelihood functions, not density functions; that is, we are not looking for the place of maximum in $\mathbf{x}$ given $\mathbf{P}$, we are looking for the place of maximum in $\mathbf{P}$ given $\mathbf{x}$.

Given an ARMA model as in equation (10.1), the residuals e_t are usually assumed to be identically distributed independent random variables with an $\mathcal{N}(0, \sigma_e^2)$ distribution, i.e., with a normal distribution of mean 0 and standard distribution σ_e, so the joint density of the residuals is

$$(10.4) \qquad \frac{1}{(2\pi)^{(n+1)/2}\sigma_e^{n+1}} \exp\left(-2^{-n-1}\sigma_e^{-2(n+1)} \sum_{t=0}^{n} \hat{e}_t^2\right).$$

Often, one likes to consider the logarithm of this, called the *log likelihood function*:

$$(10.5) \qquad -2^{-n-1}\sigma_e^{-2(n+1)} \sum_{t=0}^{n} \hat{e}_t^2 - \frac{n+1}{2}\log(2\pi) - (n+1)\log\sigma_e.$$

The the residuals $\hat{e}_t$ are functions of the parameters, i.e., of the polynomials $\phi(x)$ and $\theta(x)$ (or, rather, of their coefficients). That is, the maximum likelihood method for an ARMA model consists in finding the coefficients of these polynomials for which the unconditional sum of sqares in equation (10.3) is the least possible. That is, the maximum likelihood method in this case is a form of least squares approximation.

It is important to note that the likelihood function should not be considered a function of the residuals $\hat{e}_t$ and σ_e; properly, it is a function of the model parameters, i.e., the coefficients ϕ_k and θ_l in equation (9.1) and of the available observations of the time series $\{Y_t\}$; the residuals $\hat{e}_t$ in formulas (10.4) and (10.5) should be determined from these model parameters. The variance σ_e occurring in these formulas can also be estimated from the observations of the time series $\{Y_t\}$, but this is unimportant for the application of the maximum likelihood method.

11   The extended autocorrelation function

An ARMA(p, q) process is described by the equation

$$(11.1) \qquad \phi(B)Y_t = \theta(B)e_t,$$

where

$$\phi(x) = 1 - \sum_{k=1}^{p} \phi_k x^k \quad \text{and} \quad \theta(x) = 1 - \sum_{k=1}^{q} \theta_k x^k.$$

That is,

$$(11.2) \qquad Y_t = \sum_{k=1}^{p} \phi_k Y_{t-k} + e_t - \sum_{k=1}^{q} \theta_k e_{t-k};$$

Assuming that $\{e_t\}$ are independent normal variables of zero mean, and noting that

$$(11.3) \qquad W_t \stackrel{def}{=} Y_t - \sum_{k=1}^{p} \phi_k Y_{t-k} = e_t - \sum_{k=1}^{q} \theta_k e_{t-k},$$

we have

$$(11.4) \qquad \rho_l^{(p)} \stackrel{def}{=} \mathrm{Corr}(W_t, W_{t+l}) = 0 \quad \text{whenever} \quad l > q,$$

since on the right-hand side of W_{t+l} of (11.3) with $t + l$ replacing t, $e_{t'}$ occurs only for $t' > t$, and so $e_{t'}$ so $e_{t'}$ is independent of $Y_{t''}$ for $t'' \le t$. The quantity $\rho_l^{(p)}$ is called the *extended autocorrelation function*. A similar argument using the right-hand side of (11.3) gives

$$(11.5) \qquad \rho_l^{(p)} = \mathrm{Corr}(Y_t, W_{t+l}) = 0 \quad \text{whenever} \quad l > q,$$

11.1 The generalized Yule–Walker equations

Assuming that $\{Y_t\}$ is stationary, using the middle member of equation (11.3), equation (11.5) can be also written as

$$(11.6) \qquad \rho(l) = \sum_{k=1}^{p} \phi_k \rho(l-k) \quad \text{whenever} \quad l > q,$$

where recall that $\rho_m = \text{Corr}(Y_{t'}, Y_{t'+m})$ for any $t', m \in \mathbb{Z}$ (that is, the correlation on the right-hand side does not depend on t'). One usually considers these equations for l with $q+1 \le l \le q+p$. The equations are called the generalized Yule–Walker equations. See equation (8.4) for the Yule-Walker equation for an autoregressive process.

11.2 Determining the order of an ARMA model

In order to build an ARMA model of the correct order, one tentatively build an $\text{ARMA}(p,q)$ model, and then tests if equation (11.4) is satisfied. In building the model, given a series of observations $\{y_t\}$ of the process $\{Y_t\}$, one determines the coefficients in the $\text{ARMA}(p,q)$ model

$$Y_t = \sum_{k=1}^{p} \hat{\phi}_k Y_{t-k} + e_t - \sum_{k=1}^{q} \hat{\theta}_k e_{t-k},$$

and calculates w_t as

$$(11.7) \qquad w_t \overset{def}{=} y_t - \sum_{k=1}^{p} \phi_k y_{t-k}.$$

Then one calculates the sample autocorrelation as follows.

If y_t is available for t with $0 \le t \le n$, then w_t is available for t with $p \le t \le n-t$. One estimates the sample mean as

$$\bar{w} = \frac{1}{n+1} \sum_{i=p}^{n} w_i,$$

and then estimates the sample autocorrelation as

$$\hat{\rho}_l^{(p)} = \frac{n-p}{n-p-l} \frac{\sum_{i=p}^{n-l}(w_i - \bar{w})(w_{i+l} - \bar{w})}{\sum_{i=p}^{n}(w_i - \bar{w})^2};$$

more commonly, the first factor on the right-hand side is omitted, and one takes

$$\hat{\rho}_l^{(p)} = \frac{\sum_{i=p}^{n-l}(w_i - \bar{w})(w_{i+l} - \bar{w})}{\sum_{i=p}^{n}(w_i - \bar{w})^2}$$

instead. The difference is small, since n is usually much larger than p or l. Then one tests if $\rho_l^{(p)} = 0$ for $l > q$ (cf. equation (11.4)). For the test, one may note that the distribution of $\hat{\rho}_l^{(p)}$ is asymptotically $\mathcal{N}\big(0, 1/(n-p-l)\big)$. So the hypothesis that $\rho_l^{(p)} = 0$ is rejected with confidence of 95% if

$$|\hat{\rho}_l^{(p)}| > \frac{1.96}{\sqrt{n-p-l}} \qquad \text{for} \quad l > q.$$

If the validity of equation (11.4) is not rejected by this test, then the given $\text{ARMA}(p,q)$ is accepted as having the correct order; otherwise, and improved model needs to be built. See [12, Exhibit 6.4 on p. 117 and Exhibit 6.17 on p. 124] about how to plot the extended autocorrelation function.

12 Exponential smoothing

Various forms of exponential smoothing go back to Poisson. For time series forecasting, the two parameter exponential smoothing discussed below is usually attributed to C. C. Holt, and the three parameter version to cope with seasonality, to P. R. Winters. Assume we are given an observed time series $\{x_t\}$ $(t \geq 0)$. We want to filter out the noise to get at the core of the data. We construct a smoothed series:

$$s_0 = x_0$$
$$s_t = \alpha x_t + (1 - \alpha)s_{t-1} \qquad (t > 0),$$

where $0 < \alpha < 1$ is the smoothing factor. We will discuss how to chose α.

12.1 One-step ahead forecast

The value s_t can be used to forecast the time series $\{x_t\}$ one step ahead

$$\hat{x}_t(1) = s_t.$$

The forecasting error is

$$e_t = x_t - \hat{x}_{t-1}(1).$$

Now, if observations for t with $1 \leq t \leq n$ are available, α can be chosen by taking the sum of the squared past forecasting errors

$$\sum_{k=1}^{n} e_t^2$$

to be a minimum.

12.2 Trend: double exponential smoothing

Simple exponential smoothing does not well handle forecasting a time series with a trend. To deal with this, a trend term $\{T_t\}$ is included:

$$s_1 = x_1$$
$$T_1 = x_1 - x_0$$
$$s_t = \alpha x_t + (1 - \alpha)(x_{t-1} + T_{t-1}) \qquad (t > 2),$$
$$T_t = \beta(s_t - s_{t-1}) + (1 - \beta)T_{t-1} \qquad (t > 2),$$

where $0 < \alpha, \beta < 1$ are smoothing parameters. The h step ahead forecast will now be

$$\hat{x}_t(h) = s_t + hT_t.$$

12.3 Seasonality: triple exponential smoothing

To cope with seasonality, a seasonal term I_t and a third smoothing parameter γ with $0 < \gamma < 1$ is also introduced. The seasonal effect may be additive or multiplicative. Assume a multiplicative

seasonal effect, and consider a seasonal period s:

$$
\begin{aligned}
s_1 &= x_1 \\
s_t &= \alpha x_t + (1 - \alpha)(x_{t-1} + T_{t-1}), && (t \geq 1) \\
T_t &= \beta(s_t - s_{t-1}) + (1 - \beta)T_{t-1}, && (t \geq s) \\
I_t &= \gamma \frac{x_t}{s_t} + (1 - \gamma)I_{t-s} && (t > s).
\end{aligned}
$$

At least $2s$ steps are needed to initialize T_t and I_t. One can take

$$
T_k = \frac{1}{s^2} \sum_{i=k}^{k+s-1} (x_{i+s} - s_i) \qquad (0 \leq k < s).
$$

For the initialization of I_k one first calculates the quantities A_k; with N being the number of complete cycles present in the data $\{x_t\}$, we put

$$
A_j = \frac{1}{s} \sum_{i=1}^{s} x_{s(j-1)+i} \qquad (1 \leq j \leq N),
$$

$$
I_i = \frac{1}{s} \sum_{j=1}^{N} \frac{x_{s(j-1)+i}}{A_j} \qquad (1 \leq j \leq s).
$$

The h step ahead forecast at time t can be written as

$$
\hat{x}_t(h) = (s_t + hT_t)I_{t-s+h} \qquad (1 \leq h \leq s).
$$

For more details on exponential smoothing, see [6, §§5.2.2–5.2.8, pp. 76–80] and [19].

13 Fourier series: a brief introduction

13.1 Trigonometric series

Let f be a function on the real line. We are trying to represent f with a trigonometric series

(13.1)
$$
f(x) = \frac{1}{2}a_0 + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx);
$$

the coefficient $1/2$ in front of a_0 is used to make sure that the first equation in (13.2) below is true also in case $k = l = 0$. The series on the right-hand side is called the Fourier series of the function f. Since the trigonometric functions on the right-hand side are periodic with a period that is a multiple of 2π, for this to be possible, f must also be periodic with a period (that is a multiple) of 2π. Assuming periodicity, it is indeed possible to represent a large class of functions as a series described in formula (13.1); see Subsection 13.2 below.

The trigonometric functions satisfy the following relations, called orthogonality relations:

(13.2)
$$
\begin{aligned}
\frac{1}{\pi} \int_{-\pi}^{\pi} \cos kx \cos lx \, dx &= \begin{cases} \delta_{kl} & \text{if } l > 0, \\ 2\delta_{kl} & \text{if } l = 0, \end{cases} && (k \geq 0,\ l \geq 0), \\
\frac{1}{\pi} \int_{-\pi}^{\pi} \sin kx \sin lx \, dx &= \delta_{kl} && (k > 0,\ l > 0), \\
\frac{1}{\pi} \int_{-\pi}^{\pi} \sin kx \cos lx \, dx &= 0 && (k > 0,\ l \geq 0).
\end{aligned}
$$

These equations can easily be proved from the trigonometric formulas[13.1]

$$2\cos x \cos y = \cos(x - y) + \cos(x + y),$$
$$2\sin x \sin y = \cos(x - y) - \cos(x + y),$$
(13.3)
$$2\sin x \cos y = \sin(x + y) + \sin(x - y),$$
$$2\cos x \sin y = \sin(x + y) - \sin(x - y).$$

Ignoring issues of convergence, we multiply equation (13.1) by $\cos kx$ or by $\sin kx$ and integrate; we obtain

(13.4) $\qquad a_n = \dfrac{1}{\pi}\displaystyle\int_{-\pi}^{\pi} f(x)\cos nx\,dx \quad (n \geq 0) \quad\text{and}\quad b_n = \dfrac{1}{\pi}\displaystyle\int_{-\pi}^{\pi} f(x)\sin nx\,dx \quad (n > 0).$

13.2 Dirichlet's theorem and the Dirichlet kernel

In 1829, L. Dirichlet proved that if f is 2π-periodic and bounded, and, considered only on the interval $[-\pi, \pi)$, it has finitely many discontinuities and it is put together from finitely many monotonic pieces, then the series on the right-hand side of equation (13.1), where the coefficients are given by equations (13.4), converges to f at every point of continuity, and at a point x of discontinuity it converges to

$$\frac{1}{2}\left(\lim_{t \nearrow x} f(t) + \lim_{t \searrow x} f(t)\right).$$

Dirichlet's theorem is based on the eponymous formula[13.2] writing

(13.5) $\qquad\qquad s_n(x) = \dfrac{1}{2}a_0 + \displaystyle\sum_{k=1}^{n}(a_k \cos kx + b_k \sin kx),$

for the partial sum with the coefficients given by equations (13.4), we have

(13.6) $\qquad\qquad s_n(x) = \dfrac{1}{2\pi}\displaystyle\int_{-\pi}^{\pi} f(x - t)D_n(t)\,dt,$

where $D_n(t)$, called the Dirichlet kernel, is defined as

(13.7) $\qquad\qquad D_n(t) \overset{def}{=} 1 + 2\displaystyle\sum_{k=1}^{n}\cos kt = \dfrac{\sin\left(n + \frac{1}{2}\right)t}{\sin\frac{1}{2}t};$

for the second equation, see Problem 13.1. below.[13.3] Formula (13.6) can be proved by substituting

---

[13.1]It is convenient to list also the fourth among these equation, even though it is an easy consequence of the third one and the equation $\sin(-t) = -\sin t$.

[13.2]That is, on the formula called Dirichlet's formula.

[13.3]There are also some slightly different definitions of the Dirichlet kernel in the literature, in that some authors divide the expression in our definition by 2 or perhaps 2π.

the coefficients a_n and b_n from equations (13.4) into equation (13.5):

$$s_n(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y)\, dy + \frac{1}{\pi} \sum_{k=1}^{n} \int_{-\pi}^{\pi} f(y)(\cos ky \, \cos kx + \sin ky \, \sin kx)\, dy$$

$$(13.8) \qquad = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y)\left(1 + 2\sum_{k=1}^{n} \cos k(y-x)\right) dy = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) D_n(y-x)\, dy$$

$$= -\frac{1}{2\pi} \int_{x+\pi}^{x-\pi} f(x-t) D_n(-t)\, dt = \frac{1}{2\pi} \int_{x-\pi}^{x+\pi} f(x-t) D_n(t)\, dt$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x-t) D_n(t)\, dt;$$

here, the fourth equation was obtained by the substitution $t = x - y$, the fifth equation uses the relation $D_n(-t) = D_n(t)$, and the sixth equation uses the fact that the integrand has period 2π.

Dirichlet's formula (13.8) is the key in proving most convergence results about Fourier series, including Dirichlet's own. While such proofs involve technical difficulties of various levels that puts them beyond the scope of these notes, in the proof, first one notes that

$$(13.9) \qquad \frac{1}{2\pi} \int_{-\pi}^{\pi} D_n(t)\, dt = 1,$$

an easy consequence of the first equation in (13.7) defining the Dirichlet kernel. The key idea is that in view of formula (13.8) we have

$$f(x) - s_n(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \big(f(x) - f(x-t)\big) D_n(t)\, dt$$

$$= \frac{1}{2\pi} \left(\int_{-\eta}^{\eta} + \int_{-\pi}^{-\eta} + \int_{\eta}^{\pi}\right) \big(f(x) - f(x-t)\big) D_n(t)\, dt;$$

the first equation here holds in view of equation (13.9). The last two integrals tend to 0 as $n \to \infty$ by the Riemann-Lebesgue lemma, which says that the limits

$$\lim_{\alpha \to \infty} \int_a^b g(x) \sin \alpha x \, dx \quad \text{and} \quad \lim_{\alpha \to \infty} \int_a^b g(x) \cos \alpha x \, dx$$

are zero for any function g that is integrable in the interval $[a, b]$; this lemma is applied to the integral

$$\left(\int_{-\pi}^{-\eta} + \int_{\eta}^{\pi}\right) f(x-t) D_n(t)\, dt = \left(\int_{-\pi}^{-\eta} + \int_{\eta}^{\pi}\right) \frac{f(x-t)}{\sin \frac{1}{2}t} \sin\left(n + \frac{1}{2}\right) t \, dt$$

as $n \to \infty$. The estimation of the integral

$$\int_{-\eta}^{\eta} \big(f(x) - f(x-t)\big) D_n(t)\, dt$$

is more technical.

13.3 Problems

Problem 13.1. Prove the second equation in formula (13.7). **Hint:** expand

$$D_n(t) \sin \frac{1}{2}t$$

with the aid of the first equation in (13.7), and use the fourth equation in (13.3).

Problem 13.2. Find the Fourier series of the 2π-periodic function $f(x)$ such that $f(-\pi) = 0$ and $f(x) = x$ when $-\pi < x < \pi$.

13.4   Complex Fourier series

Using Euler's equation (3.1) and the equation obtained from it by replacing x by $-x$ and the relations $\cos(-x) = \cos x$ and $\sin(-x) = -\sin x$, we obtain the equations

$$(13.10) \qquad \cos x = \frac{e^{ix} + e^{-ix}}{2} \quad \text{and} \quad \sin x = \frac{e^{ix} - e^{-ix}}{2i}.$$

Substituting these equations into equation (13.1), we obtain the complex form of a Fourier series

$$(13.11) \qquad f(x) = \sum_{k=-\infty}^{\infty} c_k e^{ikx},$$

where

$$(13.12) \qquad c_0 = \frac{1}{2}a_0, \quad c_k = \frac{a_k - ib_k}{2}, \quad \text{and} \quad c_{-k} = \frac{a_k + ib_k}{2} \qquad (k > 0).$$

If a_k and b_k are real, we have $c_{-k} = c_k^*$, where the asterisk indicates complex conjugate.[13.4] These equations can also be written as

$$(13.13) \qquad a_0 = 2c_0, \quad a_k = c_k + c_{-k}, \quad b_k = (c_k - c_{-k})i \qquad (k > 0).$$

The orthogonality relations analogous to (13.2) can be written as

$$(13.14) \qquad \begin{aligned} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ikx} \left(e^{ilx}\right)^* dx &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ikx} e^{-ilx} \, dx \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(k-l)x} \, dx = \delta_{kl} \qquad (-\infty < k, l < \infty). \end{aligned}$$

Multiplying equation Ignoring the issues of convergence, multiplying equation (13.11) by e^{-in} and integrating, the orthogonality relations allow us to express the coefficients as

$$(13.15) \qquad c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx} \, dx \qquad (-\infty < n < \infty).$$

Again, ignoring issues of convergence, and defining the L^2 *norm* $\|f\|$ of f on the interval $[-\pi, \pi)$ by the first equation next, we obtain from equation (13.11) that

$$(13.16) \qquad \begin{aligned} \frac{1}{2\pi} \|f\|^2 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2 \, dx \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \big(f(x)\big)^* dx = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{k=-\infty}^{\infty} c_k e^{ikx} \cdot \sum_{l=-\infty}^{\infty} c_l^* e^{-ilx} \, dx \\ &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} c_k c_l^* \int_{-\pi}^{\pi} e^{i(k-l)x} \, dx = \sum_{k=-\infty}^{\infty} c_k c_k^* = \sum_{k=-\infty}^{\infty} |c_k|^2; \end{aligned}$$

[13.4] In mathematics, the complex conjugate of the number z is usually denoted by $\bar{z}$, but this notation conflicts with the notation $\bar{X}$ for the (sample) mean of the random variable X in statistics. On the other hand, it is common in mathematics to use A^* for the Hermitian conjugate of the matrix A, and in a way this is an analog of the complex conjugate for matrices.

the penultimate[13.5] equation on above holds because the orthogonality relations. The square root of sum on the right is called the l^2 norm of the sequence $\{c_k\}_{k=-\infty}^{\infty}$. The equation of the left- and right-hand sides in this formula is called Parseval's identity, named after Marc-Antoine Parseval. He claimed this identity without proof as self-evident in a paper dated 1799. A similar calculation involving equation (13.1) gives Parseval's formula for the coefficients a_k and b_k

$$(13.17) \qquad \frac{1}{\pi}\|f\|^2 = \frac{1}{\pi}\int_{-\pi}^{\pi}\big(f(x)\big)^2\,dx = \frac{|a_0|^2}{2} + \sum_{k=1}^{\infty}\big(|a_k|^2 + |b_k|^2\big).$$

One usually considers this equation only for real a_k and b_k, in which case the absolute values can be omitted.

In undergraduate courses, when integration is defined precisely, usually the integral concept introduced by Bernhard Riemann in 1854 is discussed in his Habilitationsschrift. [13.6] For a deeper understanding of the L^2 norm a newer integral concept, introduced by Henri Lebesgue in 1904 is needed. We will not go into these issues.

13.5   Problem

Problem 13.3. Use the solution of Problem 13.2 and Parseval's formula (13.17) to prove

$$(13.18) \qquad \sum_{n=1}^{\infty}\frac{1}{n^2} = \frac{\pi^2}{6},$$

a formula first proved by Euler in 1741.[13.7]

13.6   The complex form of the Dirichlet kernel

The Dirichlet kernel defined in equation (13.7) can also be written as

$$
\begin{aligned}
D_n(t) &= \sum_{k=-n}^{n} e^{ikx} = e^{-inx}\,\frac{e^{i(2n+1)x}-1}{e^{ix}-1} = \frac{e^{i(n+1)x}-e^{-inx}}{e^{ix}-1} \\
&= \frac{e^{i(n+1/2)x}-e^{-i(n+1/2)x}}{e^{ix/2}-e^{-ix/2}} = \frac{\sin\big(n+\frac{1}{2}\big)t}{\sin\frac{1}{2}t};
\end{aligned}
$$

(13.19)

the second equation is obtained by using the sum formula for the geometric progression, and the last one follows from the second one among Euler's equations (13.10). The symmetric partial sum

$$(13.20) \qquad s_n(x) = \sum_{k=-n}^{n} c_k e^{ikx} \qquad (n \geq 0),$$

where the coefficients are given by equations (13.15), is given by equation (13.8).

[13.5]The one before the last.

[13.6]Habilitation is a post-doctoral qualification at universities in various countries; there is a Wikipedia article about this. Habilitationsschrift is the name of the dissertation used for habilitation in German speaking universities. There is a German Wikipedia article about this, but it is written in German.

[13.7]The problem of finding the sum of the series on the left is called the Basel problem, posed by Pietro Mengoli in 1644, and solved by Euler in 1734, though he was not able to justify his arguments rigorously until 1741.

14 Trigonometric interpolation with equidistant nodes

14.1 Lagrange interpolation

Let N be a positive integer, let z_1, z_2, ...,z_N be distinct complex numbers, and let w_1, w_2, ...,w_N also be complex numbers, these latter not necessarily distinct. The task of *polynomial interpolation* is to find a polynomial $P(z)$ of degree less than N such that $P(z_k) = w_k$ for k with $1 \leq k \leq N$; the points z_k are called *interpolation points* or *nodes*. It is not hard to prove that if there is such a polynomial, then it is unique; see Problem 14.1 below. We will show that there is indeed a polynomial $P(z)$ of degree less than N satisfying these requirements. The polynomial interpolation problem was first solved by Newton. A different, elegant solution was later found by Lagrange. Here we consider the latter solution.

Lagrange considers the polynomials

$$l_k(z) = \prod_{\substack{j=1 \\ j \neq k}}^{N} \frac{z - z_j}{z_k - z_j}.$$

It is clear that $l_k(z)$ is a polynomial of degree $N - 1$, since the numbers in the denominator do not depend on z. Further, for any integer j with $1 \leq j \leq N$ we have

$$l_k(z_j) = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k. \end{cases}$$

Indeed, if $z = z_k$ then each of the fractions in the product expressing $l_k(z)$ is 1, and if $z = z_j$ for $j \neq k$ then one of the fractions in this product has a zero numerator. For this reason, the polynomial $P(z)$ defined as

$$P(z) = \sum_{k=1}^{N} w_k l_k(z)$$

satisfies the requirements; that is $P(z_k) = w_k$ for k with $1 \leq k \leq N$.

Both Lagrange's and Newton's solution of the polynomial interpolation problem has uses. Only Newton's solution is suitable for numerical calculations; both solutions have theoretical applications. It is easy to see that Lagrange's solution is works for any field replacing the field of complex numbers.

14.2 Problem

Problem 14.1. Show that the solution of the interpolation problem is unique. That is, given points z_1, z_2, ...,z_N be distinct complex numbers, and w_1, w_2, ...,w_N complex numbers, these latter not necessarily distinct. Let $P_1(z)$ and $P_2(z)$ be polynomials of degree less than N such that $P_1(z_k) = P_2(z_k) = w_k$ for all k with $1 \leq k \leq N$. Show that then $P_1(z)$ and $P_2(z)$ are the same polynomial.

14.3 Complex exponential interpolation with equidistant nodes

Let f be a 2π-periodic function on the real line, and let x_0 be an arbitrary fixed real. We want to represent f at the nodes

(14.1) $$x_n = x_0 + 2n\pi/N \qquad (0 \leq n < N)$$

by a complex exponential polynomial as

$$(14.2) \qquad f(x_n) = \sum_{k=0}^{N-1} c_k e^{ikx_n} \qquad (0 \leq n < N).$$

It is immediate from the main result of Subsection 14.1 that this problem has a solution. Namely, the question can be reformulated as the polynomial interpolation problem of finding a polynomial

$$P(z) = \sum_{k=0}^{N-1} c_k z^k$$

such that $P(e^{ix_n}) = f(x_n)$ for n with $0 \leq n < N$. [14.1] To find the coefficients c_k in (14.2), observe that for k with $0 < k < N$ we have

$$\sum_{n=0}^{N-1} e^{2nk\pi i/N} = \frac{e^{2kN\pi i/N} - 1}{e^{2k\pi i/N} - 1} = \frac{e^{2k\pi i} - 1}{e^{2k\pi i/N} - 1} = 0;$$

the first equation holds by the sum formula of geometric series, and the second equation holds since $e^{2k\pi i} = 1$; this calculation is not applicable in case $n = 0$ since the denominator is 0 then.[14.2] For $n = 0$, all terms in the sum are 1, and hence the sum is N. Thus

$$(14.3) \qquad \begin{aligned} \frac{1}{N} \sum_{n=0}^{N-1} e^{x_n k i} &= \frac{1}{N} \sum_{n=0}^{N-1} e^{k(x_0 + 2n\pi/N)i} \\ &= \frac{1}{N} e^{ikx_0} \sum_{n=0}^{N-1} e^{2nk\pi i/N} = \begin{cases} 1 & \text{if } k = 0, \\ 0 & \text{if } k \neq 0 \quad (-N < k < N). \end{cases} \end{aligned}$$

Hence, the orthogonality relations are

$$(14.4) \qquad \frac{1}{N} \sum_{n=0}^{N-1} e^{ikx_n} e^{-ilx_n} = \frac{1}{N} \sum_{n=0}^{N-1} e^{i(k-l)x_n} = \delta_{kl} \qquad (0 \leq k, l < N).$$

Multiplying equation (14.2) by e^{-ilx_n} and summing for n, by using these orthogonality relations we obtain

$$(14.5) \qquad c_l = \frac{1}{N} \sum_{n=0}^{N-1} f(x_n) e^{-ilx_n} \qquad (0 \leq l < N).$$

Observe that the right-hand side here can be considered as a numerical integration formula approximating the integral in (13.15) (with l here replacing n in that formula). Parseval's identity can be written as

$$(14.6) \qquad \frac{1}{N} \sum_{n=0}^{N-1} |f(x_n)|^2 = \sum_{n=0}^{N-1} |c_n|^2.$$

[14.1] Actually, this equation holds for all n, not just for n with $0 \leq n < N$, but this is a consequence of the equations $f(ix_{n+N}) = f(ix_n)$ and $P(e^{ix_{n+N}}) = P(e^{ix_n})$.

[14.2] More generally, the calculation is not applicable for any k that is divisible by N, since for such k the denominator is 0; it is applicable for any other $k \in \mathbb{Z}$. That is, the above formula is true for any integer k that is not divisible by N.

14.4 More on complex exponential interpolation with equidistant nodes

Instead of the range $0 \leq k < N$ of exponent range in equation (14.2) we can take a exponent range $K \leq k < K + N$ for an arbitrary $K \in \mathbb{Z}$. Finding a complex exponential polynomial

$$(14.7) \qquad Q_K(x) = \sum_{k=K}^{K+N-1} c_k e^{ikx}$$

such that

$$(14.8) \qquad Q_K(x_n) = f(x_n)$$

for n with $0 \leq n < N$ is equivalent to the type of problem given in equation (14.2); namely, we need to find a complex exponential polynomial

$$Q(x) = \sum_{k=0}^{N-1} c_{k+K} e^{ikx}$$

such that $Q(x_n) = f(x_n)e^{-iKx_n}$ for for n with $0 \leq n < N$. As for the coefficients c_k, multiplying equation (14.8) with $x = x_n$ by e^{-lx_n} and summing for n, the orthogonality relations (14.4) give the equation

$$(14.9) \qquad c_l = \frac{1}{N} \sum_{n=0}^{N-1} f(x_n) e^{-ilx_n} \qquad (K \leq l < K + N).$$

This is of course the same as equation (14.5), except that a different range of the coefficients c_k is considered. Parseval's identity to replace equation (14.6) can now be written as

$$(14.10) \qquad \frac{1}{N} \sum_{n=0}^{N-1} |f(x_n)|^2 = \sum_{n=K}^{K+N-1} |c_n|^2.$$

14.5 Real trigonometric interpolation with an odd number of equidistant nodes

Translating these formulas to the real line is easier in case N is odd. Assuming this and writing $N = 2M + 1$, consider equation (14.7) with $K = -M$. Assuming $f(x)$ is real for all $x \in \mathbb{R}$, equation (14.9) implies that $c_{-l} = c_l*$ for l with $-M \leq l \leq M$. Hence, Euler's equations (13.10), we obtain

$$(14.11) \qquad Q_{-M}(x) = \sum_{k=-M}^{M} c_k e^{ikx} = \frac{a_0}{2} + \sum_{k=1}^{M} (a_k \cos kx + b_k \sin kx)$$

with

$$(14.12) \qquad a_0 = 2c_0, \quad a_k = c_k + c_{-k} = 2\Re c_k, \quad b_k = (c_k - c_{-k})i = -2\Im c_k \qquad (0 < k \leq M);$$

the second equations for a_k and b_k hold since $c_{-k} = c_k^*$. These equations identical to equations (13.13). These equations together with the Euler equations (13.10) and equation (14.9) give the

equations for the coefficients:

$$a_k = \frac{2}{N} \sum_{n=0}^{N-1} f(x_n) \cos kx_n \qquad (0 \leq k \leq M),$$

$$b_k = \frac{2}{N} \sum_{n=0}^{N-1} f(x_n) \sin kx_n \qquad (0 < k \leq M).$$

That is, according to equation (14.8) we have

(14.14) $$f(x_n) = \frac{a_0}{2} + \sum_{k=1}^{M} (a_k \cos kx_n + b_k \sin kx_n) \qquad (0 \leq n < N, \ N = 2M+1).$$

with these coefficients. As it can be seen, the number of coefficients here is $2M + 1 = N$. The coefficient equations can also be obtained directly from the orthogonality relations

(14.15)

$$\frac{2}{N} \sum_{n=0}^{N-1} \cos kx_n \cos lx_n = \begin{cases} \delta_{kl} & \text{if } l > 0, \\ 2\delta_{kl} & \text{if } l = 0, \end{cases} \qquad (0 \leq k, l \leq M),$$

$$\frac{2}{N} \sum_{n=0}^{N-1} \sin kx_n \sin lx_n = \delta_{kl} \qquad (1 \leq k, l \leq M),$$

$$\frac{2}{N} \sum_{n=0}^{N-1} \sin kx_n \cos lx_n = 0 \qquad (1 \leq k \leq M, \ 0 \leq l \leq M).$$

To prove these equations, we need first observe that

(14.16) $$\sum_{n=0}^{N-1} \sin(kx_n + \alpha) = \sum_{n=0}^{N-1} \cos(kx_n + \alpha) = 0 \qquad (1 \leq |k| < N \text{ and } \alpha \in \mathbb{R}).$$

These equations are valid for both even and odd N. They follow by multiplying (14.3) by $e^{i\alpha}$ and taking real parts and imaginary parts, respectively. Then equations (14.15) can be easily proved using the trigonometric formulas (13.3). Parseval's equation can be written in this case as

(14.17) $$\frac{2}{N} \sum_{n=0}^{N-1} |f(x_n)|^2 = \frac{|a_0|^2}{2} + \sum_{n=1}^{M} (|a_n|^2 + |b_n|^2).$$

14.6 Real trigonometric interpolation with an even number of equidistant nodes

The case of even $N > 0$ is somewhat more complicated. Given a real-valued 2π-periodic function f on $\mathbb{R}$, with x_n as in equation (14.1), writing $N = 2M$, we require

(14.18) $$f(x_n) = \frac{a_0}{2} + \sum_{k=1}^{M-1} (a_k \cos kx_n + b_k \sin kx_n) + \frac{a_M}{2} \cos\big(M(x_n - x_0)\big) \qquad (0 \leq n < N).$$

The number of coefficients here is also N. This equation can be justified as follows. Represent f with the interpolation formula in (14.7) with $K = -M + 1$; that is, also using equation (14.7), we

have

$$(14.19) \qquad f(x_n) = \sum_{k=-M+1}^{M} c_k e^{ikx_n}.$$

According to equation (14.1), we have $M(x_n - x_0) = n\pi$, and so equation (14.9) implies

$$(14.20)$$
$$c_M e^{iMx_0} = \frac{1}{N} \sum_{n=0}^{N-1} f(x_n) e^{-iM(x_n-x_0)} = \frac{1}{N} \sum_{n=0}^{N-1} f(x_n) e^{-in\pi} = \frac{1}{N} \sum_{n=0}^{N-1} f(x_n)(-1)^n$$
$$= \frac{1}{N} \sum_{n=0}^{N-1} f(x_n) \cos n\pi = \frac{1}{N} \sum_{n=0}^{N-1} f(x_n) \cos\big(M(x_n - x_0)\big).$$

Equations (14.12) are now replaced with

$$(14.21) \qquad \begin{aligned} a_0 &= 2c_0, \quad a_M = 2c_M e^{iMx_0}, \\ a_k &= c_k + c_{-k} = 2\Re c_k, \quad b_k = (c_k - c_{-k})i = -2\Im c_k \qquad (0 < k < M). \end{aligned}$$

all these equations except for the second one can be justified the same way as in formula (14.12). The reason for the second equation will be clear soon.

It is easy to see that with this choice of the coefficients a_k and b_k we can define a trigonometric polynomial $\tilde{f}(x)$ such that $\tilde{f}(x_n) = f(x_n)$, as required in equation (14.18):

$$(14.22)$$
$$\tilde{f}(x) \stackrel{def}{=} \frac{a_0}{2} + \sum_{k=1}^{M-1} (a_k \cos kx + b_k \sin kx) + \frac{a_M}{2} \cos\big(M(x - x_0)\big)$$
$$= \sum_{k=-M+1}^{M-1} c_k e^{ikx} + \Re(c_M e^{iMx});$$

here the first equation defines the interpolation polynomial in case of an even number of nodes; the second equation here needs some explanation. By virtue of the second equation in (14.21) we have

$$\Re(c_M e^{iMx}) = \Re\left(\frac{a_M}{2} e^{iM(x-x_0)}\right) = \frac{a_M}{2} \cos(M(x - x_0));$$

the last equation holds since $a_M = 2c_M e^{iMx_0}$ is real according to formula (14.20). The second equation in (14.22) follows from this equation. This also justifies the adoption of the second equation in (14.21). Further, the observation just made that $a_M = 2c_M e^{iMx_0}$ is real also implies that $c_M e^{iMx_n}$ is real for every integer n. Indeed, we have

$$c_M e^{iMx_n} = c_M e^{ix_0} e^{iM(x_n-x_0)} = c_M e^{ix_0} e^{in\pi} = c_M e^{ix_0}(-1)^n;$$

the third equation here holds in view of the definition of x_n given in formula (14.1). Thus, for $x = x_n$ the right-hand side of (14.22) equals

$$\tilde{f}(x_n) = \sum_{k=-M+1}^{M-1} c_k e^{ikx_n} + \Re(c_M e^{iMx_n}) = \sum_{k=-M+1}^{M-1} c_k e^{ikx_n} + c_M e^{iMx_n}$$
$$= \sum_{k=-M+1}^{M} c_k e^{ikx_n} = f(x_n);$$

the real part $\Re(\cdot)$ was dropped from the second member of these equation, since the term it was applied to is real. The last equality holds according to equation (14.19). The equality of the sides show that the trigonometric polynomial $\tilde{f}(x)$ indeed interpolates $f(x)$ at the given notes, justifying the definition of the interpolation polynomial in equation (14.22).

The orthogonality relations corresponding to the above equations are

$$
(14.23) \quad
\begin{aligned}
\frac{2}{N} \sum_{n=0}^{N-1} \cos k x_n \cos l x_n &= \begin{cases} \delta_{kl} & \text{if } l > 0, \\ 2\delta_{kl} & \text{if } l = 0, \end{cases} && (0 \le k, l < M), \\[2mm]
\frac{2}{N} \sum_{n=0}^{N-1} \cos k x_n \cos\big(M(x_n - x_0)\big) &= 0 && (0 \le k < M), \\[2mm]
\frac{2}{N} \sum_{n=0}^{N-1} \cos^2 M(x_n - x_0) &= 2, \\[2mm]
\frac{2}{N} \sum_{n=0}^{N-1} \sin k x_n \sin l x_n &= \delta_{kl} && (1 \le k, l \le M), \\[2mm]
\frac{2}{N} \sum_{n=0}^{N-1} \sin k x_n \cos l x_n &= 0 && (1 \le k, l < M), \\[2mm]
\frac{2}{N} \sum_{n=0}^{N-1} \sin k x_n \cos\big(M(x_n - x_0)\big) &= 0 && (1 \le k < M),
\end{aligned}
$$

The third of these equations holds since $\cos\big(M(x_n - x_0)\big) = \cos n\pi = (-1)^n$, as we saw in equation (14.20), the rest follows the same way as (14.15) from the equations in (14.16) and the trigonometric formulas (13.3). Using the orthogonality relations, equation (14.18) implies that

$$
(14.24) \quad
\begin{aligned}
a_k &= \frac{2}{N} \sum_{n=0}^{N-1} f(x_n) \cos k x_n && (0 \le k < M), \\[2mm]
a_M &= \frac{2}{N} \sum_{n=0}^{N-1} f(x_n) \cos M(x_n - x_0), \\[2mm]
b_k &= \frac{2}{N} \sum_{n=0}^{N-1} f(x_n) \sin k x_n && (0 < k < M).
\end{aligned}
$$

Parseval's equation becomes

$$
(14.25) \quad \frac{2}{N} \sum_{n=0}^{N-1} |f(x_n)|^2 = \frac{|a_0|^2}{2} + \sum_{n=1}^{M-1} \big(|a_n|^2 + |b_n|^2\big) + \frac{|a_M|^2}{2}.
$$

15 The Stieltjes integral

There is a clear analogy between the formulas describing Fourier series and trigonometric interpolation with equidistant nodes. This analogy can be brought out more clearly by rewriting the interpolation formulas with the aid of Stieltjes integrals. The next three definitions describe the

Riemann–Stieltjes integral.[15.1]

Definition 15.1 (Partition). A *partition* of the interval $[a, b]$ is a finite sequence $\langle x_0, x_1, \ldots, x_n \rangle$ of points such that

$$P : a = x_0 < x_1 < x_2 < \ldots < x_n = b.$$

The *width* or *norm* of a partition is

$$\|P\| \stackrel{def}{=} \max\{x_i - x_{i-1} : 1 \le i \le n\}.$$

Definition 15.2 (Riemann–Stieltjes sum). Given a partition

$$P : a = x_0 < x_1 < x_2 < \ldots < x_n = b.$$

of the interval $[a, b]$, a *tag* for the interval $[x_{i-1}, x_i]$ with $1 \le i \le n$ is a number $\xi_i \in [x_{i-1}, x_i]$ for each i. A partition with a tag for each interval $[x_{i-1}, x_i]$ is called a *tagged* partition. Given a tagged partition as described, and given the functions f and g on $[a, b]$, the corresponding Riemann–Stieltjes sum is

$$S = \sum_{i=1}^{n} f(\xi_i)\big(g(x_i) - g(x_{i-1})\big).$$

The Riemann–Stieltjes integral

$$\int_a^b f(x)\, dg(x)$$

is defined as the limit of the Riemann–Stieltjes sums S associated with the partition P as $\|P\| \to 0$, independently of the choice of the tags. While not important for our purposes, we will give a rigorous definition (if we take $g(x) = x$ in this definition, this gives the usual definition of the Riemann inegral):

Definition 15.3 (Riemann–Stieltjes integral). If there is a real number A such that for every $\epsilon > 0$ there is a $\delta > 0$ such that for any Riemann–Stieltjes sum S for f and g associated with a partition of width $< \delta$ of $[a, b]$ we have $|A - S| < \epsilon$, then we call A the *Riemann–Stieltjes integral* of f with respect to g on $[a, b]$, and we write $A = \int_a^b f\, dg$. In this case we call f *Riemann–Stieltjes integrable* with respect to g on $[a, b]$.

Let $N > 0$, and let the interpolation points x_n be chosen as formula (14.1). except now we want to consider x_n for any $n \in \mathbb{Z}$. That is, given some real x_0

(15.1) $$x_n = x_0 + 2n\pi/N \qquad (n \in \mathbb{Z}).$$

Assume f is a continuous 2π-periodic function, and define the function ω_N on $\mathbb{R}$ as

(15.2) $$\omega_N(x) = \frac{2\pi n}{N} \quad \text{if} \quad x_n \le x < x_{n+1} \qquad (n \in \mathbb{Z}).$$

Then we can write equation (14.5) as

$$c_l = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ilx}\, d\omega_N(x) \qquad (0 \le l < N),$$

[15.1]There is also a Lebesgue–Stieltjes integral that extends the concept of of Riemann–Stieltjes integrability. See footnote 13.6 on p. 46. The x_i in the definition that follows has nothing to do with the nodes of the interpolation also denoted by x_i in a different context.

emphasizing the analogy with equation (13.15) – cf. Problem 15.1. Since we assumed that f is 2π-periodic, we could integrate on any interval of length 2π instead of $[-\pi, \pi]$.[15.2] Other equations involving interpolations can also be rewritten as Stieltjes integrals. In case $N = 2M + 1$ is odd, writing, in analogy with formula (13.8),

$$(15.3) \qquad \tilde{f}(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) D_M(x - y)\, d\omega_N(y),$$

then we have

$$(15.4) \qquad \tilde{f}(x_n) = f(x_n) \qquad \text{for} \quad 0 \le n < N.$$

The easiest way to see this is use real form of interpolation as given in equation (14.11), and repeating the calculations in equation (13.8) with M replacing n; see Problem 15.2 below. For even N, equation (13.8) needs a minor modification.[15.3]

15.1   More on Stieltjes integrals

The only reason we mentioned Stieltjes integrals is to more closely highlight the analogy between Fourier series and trigonometric interpolation. We will include here some simple results to put Stieltjes integrals in the proper context, even though they are not needed for the discussion below. The first one converts Stieltjes integrals into Riemann integrals in certain cases (but not in the case of interest to us above, when the function playing the role of g is not continuous).

Theorem 15.1. *Assume g is differentiable on $[a, b]$. Assume further that the Riemann integral $\int_a^b f(x)g'(x)\, dx$ and the Riemann–Stieltjes integral $\int_a^b f(x)\, dg(x)$ exist. Then*

$$\int_a^b f(x)\, dg(x) = \int_a^b f(x)g'(x)\, dx.$$

Proof. Let
$$P : a = x_0 < x_1 < x_2 < \ldots < x_n = b.$$
a partition of the interval $[a, b]$. By the mean-value theorem of differentiation, for each i with $1 \le i \le n$ there is a $\xi_i \in [x_{i-1}, x_i]$ such that $g'(\xi_i)(x_i - x_{i-1}) = g(x_i) - g(x_{i-1})$.[15.4] Hence we have

$$\sum_{i=1}^n f(\xi_i)\big(g(x_i) - g(x_{i-1})\big) = \sum_{i=1}^n f(\xi_i)g'(\xi_i)(x_i - x_{i-1}).$$

Making $\|P\| \to 0$, the left-hand side tends to $\int_a^b f(x)\, dg(x)$ and the right-hand side tends to $\int_a^b f(x)g'(x)\, dx$, completing the proof. $\qquad\square$

Theorem 15.2 (Integration by Parts). *Assume the integral $\int_a^b f(x)\, dg(x)$ is defined. Then the integral $\int_a^b g(x)\, df(x)$ is also defined and we have*

$$\int_a^b f(x)\, dg(x) = f(b)g(b) - f(a)g(a) - \int_a^b g(x)\, df(x).$$

[15.2]If we integrate on the interval $[x_m, x_m + 2\pi]$, then the value of the integrand at $x_{n+N} = x_m + 2\pi$ will contribute to the integral, and its value an x_m will not. This is because in equation (15.2) we defined $\omega_N(x)$ to be constant on the interval $[x_n, x_{n+1})$.

[15.3]In formula (15.3) we used the fourth member of formula (13.8), since the rest of the transformations in that formula are not valid in the present situation.

[15.4]In order to apply the mean-value theorem, we need to assume that g is real valued, since the mean-value theorem is not valid for complex-valued functions. The result can nevertheless be proved in case g is complex valued by establishing it separately for the real and the imaginary parts of g.

Proof. For the proof, we redefine the concept of partition by allowing $P = \langle x_i : 1 \leq i \leq n \rangle$ to be a nondecreasing sequence. This is a harmless change, since the terms $f(\xi_i)\big(g(x_i) - g(x_{i-1})\big)$ for which $x_{i-1} = x_i$ do not contribute to the Riemann–Stieltjes sum. Let P be such an arbitrary partition; that is

$$P : a = x_0 \leq x_1 \leq x_2 \leq \ldots \leq x_n = b,$$

and let $\xi_i \in [x_{i-1}, x_i]$ be arbitrary tags. We have the identity

$$\sum_{i=1}^{n} f(\xi_i)\big(g(x_i) - g(x_{i-1})\big) = f(x_n)g(\xi_n) - f(\xi_1)g(x_0) + \sum_{i=1}^{n-1} g(x_i)\big(f(\xi_i) - f(\xi_{i+1})\big).$$

This is easy to verify; namely, the same terms are added on both sides, in different order. Indeed, for i with $1 \leq i \leq n$, both sides adds the term $+f(\xi_i)g(x_i)$, except that on the right-hand time for $i = n$ this term is written out separately. Further, both sides adds the terms $-f(\xi_i)g(x_{i-1})$ for i with $1 \leq i \leq n$, even though on the right-hand side this term is written as $-f(\xi_{i+1})g(x_i)$ for i with $1 \leq i \leq n - 1$, and the term corresponding to $i = 0$, i.e., the term $-f(\xi_1)g(x_0)$, is written out separately. This rearrangement of a sum is called *partial summation* or *Abel rearrangement*, named after the Norwegian mathematician Niels Henrik Abel.

Making the assumption $a = x_0 = \xi_1 = x_1$ and $x_{n-1} = \xi_n = x_n = b$, the above identity becomes

$$\sum_{i=1}^{n} f(\xi_i)\big(g(x_i) - g(x_{i-1})\big) = f(b)g(b) - f(a)g(a) - \sum_{i=1}^{n-1} g(x_i)\big(f(\xi_{i+1}) - f(\xi_i)\big).$$

Considering

$$P' : a = \xi_1 \leq \xi_2 \leq \xi_3 \leq \ldots \leq \xi_n = b$$

with the tags $x_i \in [\xi_i, \xi_{i+1}]$ for i with $1 \leq i \leq n - 1$, the right-hand side contains a Riemann–Stieltjes sum for the integral $\int_a^b g(x)\, df(x)$, and the left-hand side contains a Riemann–Stieltjes sum for the integral $\int_a^b f(x)\, dg(x)$; the fact that we allow $x_{i-1} = x_i$ makes no difference here, since the terms with $x_{i-1} = x_i$ make no contribution to the sum.[15.5] Since $\xi_{i-1} \leq x_i \leq \xi_i \leq x_{i+1} \leq \xi_{i+1}$ for all i with $1 \leq i \leq n - 1$, $x_0 = x_1$, and $x_{n-1} = x_n$, we have $\|P\| \leq 2\|P'\|$. Hence, making $\|P'\| \to 0$, we also have $\|P\| \to 0$; hence the left-hand side tends to $\int_a^b f(x)\, dg(x)$, since this integral was assumed to exist. So, the right-hand side also has a limit; thus, the integral $\int_a^b g(x)\, df(x)$ also exists, it being the limit of the sum on the right-hand side. This completes the proof of the theorem. $\qquad\square$

We also have a change of variables (i.e., substitution) formula for Riemann–Stieltjes integrals; it is even simpler than the one for regular Riemann integrals. For this, we need to put

$$\int_b^a f(x)\, dg(x) \stackrel{def}{=} -\int_a^b f(x)\, dg(x) \qquad (a < b),$$

as is usual in case of Riemann integrals. At this point, we might as well put $\int_a^b f(x)\, dg(x) = 0$ in case $a = b$.

Theorem 15.3. *Assume the integral $\int_a^b f(x)\, dg(x)$ exists, and let $h : [A, B] \to [a, b]$ be a nondecreasing or nonincreasing function onto $[a, b]$. Then the integral $\int_A^B f\big(h(t)\big)\, dg\big(h(t)\big)$, exists and we have*

$$\int_A^B f\big(h(t)\big)\, dg\big(h(t)\big) = \int_{h(A)}^{h(B)} f(x)\, dg(x).$$

Note that $h(A) = a$ and $h(B) = b$ in case h is nondecreasing, and $h(A) = b$ and $h(B) = a$ in case h is nonincreasing. As for the proof, it is fairly direct and straightforward except that it involves simple results about uniform continuity, and so we omit the proof.[15.6] Readers familiar with uniform continuity can easily construct a proof.

[15.5] The equality $\xi_i = \xi_{i+1}$ is possible, whether or not we allow the possibility that $x_{i-1} = x_i$. This causes no trouble, just as allowing $x_{i-1} = x_i$ causes no trouble.

[15.6] A function h satisfying the requirements of Theorem 15.3 is necessarily continuous, and so also uniformly continuous.

15.2 Problems

Problem 15.1. Let f be a function on $[-1, 1]$ that is continuous at 0, and let g be the function that is

$$g(x) = \begin{cases} 0 & \text{if } -1 \le x < 0, \\ 1 & \text{if } 0 \le x \le 1. \end{cases}$$

Show that

$$\int_{-1}^{1} f(x) \, dg(x) = f(0).$$

Problem 15.2. Re-write formula (15.3) with a sum instead as a Stieltjes integral.

Problem 15.3. Prove equation (15.4).

Problem 15.4. Write the equation corresponding the equations (15.3) and (15.4) with an even number of points. **Hint:** the formula you obtain is essentially identical, but with a slightly modified version of the Dirichlet kernel.

16 Spectrum of a time series

Given the observations y_t of a time series for $0 \le t < N$ we want to represent it in the form

$$(16.1) \qquad y_t = \sum_{k=0}^{N-1} c_k e^{2ikt\pi/N} \qquad (0 \le t < N).$$

This is just the interpolation formula (14.2) with $y_t = f(x_t)$ and $x_t = 2t\pi/N$; other forms of the interpolation formula could also have been used instead. The term $c_k e^{2ikt\pi/N}$ is said to represent the frequency k/N in the above sum. That is, y_t is decomposed as a sum of frequencies.

Given a complex number z, it can be written in what is called a trigonometric form $z = |z|e^{i\alpha}$; here α is called the argument of z, and it is determined only up to an additive multiple of 2π (since $e^{2\pi i} = 1$). We write $\alpha = \arg z$; we usually take $0 \le \arg z < 2\pi$, though occasionally other values of $\arg z$ may be taken. According to (14.5), we have

$$(16.2) \qquad c_k = \frac{1}{N} \sum_{t=0}^{N-1} y_t e^{-2ikt\pi/N} \qquad (0 \le k < N).$$

Here $|c_k|$ is called the amplitude of the frequency k/N and $\arg c_k$, its phase in the time series y_t, $0 \le t < N$.

16.1 The periodogram

In equation (16.1), the frequency ranges from 0 to $(N-1)/N$, so roughly from 0 to 1. We will assume that y_t is real. Hence we have $c_{N-k} = c_k^*$; this is clear from equation (16.2), since

$$e^{-2i(N-k)t\pi/N} = e^{-2iNt\pi/N} \cdot e^{2ikt\pi/N} = e^{2ikt\pi/N} = \left(e^{-2ikt\pi/N} \right)^*.$$

For this reason, $|c_k|$ for $k > N/2$, is of no interest. That is, the amplitude is only of interest in the range $[0, 1/2]$.[16.1]

[16.1] In the real interpolation formula (14.14) or (14.15), the frequencies clearly range from 0 to 1/2, but much of the discussion is simpler with the complex interpolation formula (14.2).

In electromagnetic radiation, the squares of the amplitudes of the electric and magnetic fields are proportional to the energy density of the the radiation. Analogously, for $1 \leq k \leq N/2$, the term $c_k e^{2ikt\pi/N}$ contributes an energy of $|c_k|^2$. This energy is restricted to the frequency range

$$\left(\frac{2k-1}{2N}, \frac{2k+1}{2N} \right).$$

Given that the length of this interval is $1/N$, this gives an energy density of $N|c_k|^2$. Given that t is an integer, we have

$$c_{N-k} e^{2i(N-k)t\pi/N} = c_{N-k} e^{-2ikt\pi/N}.$$

hence this term represents the same frequency,[16.2] and it contributes an additional energy of $|c_{N-k}|^2$ to the same frequency range. Since we have $c_{N-k} = c_k^*$; we have $|c_{N-k}|^2 = |c_k|^2$. This makes the energy density at the frequency k/n to be $2N|c_k|^2$ for $1 \leq k < N/2$. If N is even and $k = N/2$, then $N - k = k$, so only one term contributes to the energy density. Thus we define the *periodogram* as

(16.3)
$$I\left(\frac{k}{N} \right) \overset{def}{=} 2N|c_k|^2 \qquad (1 \leq k < N/2),$$

$$I\left(\frac{1}{2} \right) \overset{def}{=} N|c_{N/2}|^2 \quad \text{if } N \text{ is even.}$$

With this notation, equation (14.6) becomes

$$\sum_{t=0}^{N-1} y_t^2 - N|c_0|^2 = N \sum_{n=1}^{N-1} |c_n|^2 = \sum_{n=1}^{\lfloor N/2 \rfloor} I\left(\frac{k}{N} \right).$$

The term $|c_0|^2$ does not represent a wave, and so it carries no "energy."

16.2 Sampling rate and the Nyquist frequency

We saw that the time series above can be described in terms of frequencies in the range $[0, 1/2]$. In electric engineering, one has a continuous time series (Voltage, for example), and one takes measurements of this time series, to represent the continuous time series as a discrete[16.3] time series, perhaps for digital transmission or recording. Often, one wants to reconstruct the continuous time series. This is the situation, for example, with the transmission or digital recording of sound. Sound is really an oscillation at various frequencies, and an accurate reconstruction of these frequencies is important. The considerations above show that only frequencies in the range $[0, 1/2]$ can be reconstructed, where the unit time is the time between samples. That is, to reconstruct sound waves in the range of 0 to 6000 Hz[16.4] one needs to sample the signal representing the sound 12000 times a second. That is, the sampling rate must be twice the maximum frequency that can be reconstructed from the signal. This maximum frequency is called the Nyquist frequency after the Swedish-born American electronic engineer Harry Nyquist.

[16.2]Unless but one wants to consider negative frequencies; however, in the real form (14.14) there are no negative frequencies.

[16.3]It is important to learn the difference between *discrete* and *discreet*.

[16.4]Cycles per second. Named after the German physicist Heinrich Rudolf Hertz, who experimentally demonstrated the existence of electromagnetic waves, after the theory of electromagnetism developed by James Clerk Maxwell predicted their existence.

16.3   Variance of a complex-valued random variable

Below, we are going to calculate the variance of a complex-valued random variable. If X is a complex-valued random variable, we define its variance as

$$\mathrm{Var}(X) = \mathrm{E}\big(|X - \mathrm{E}(X)|^2\big).$$

We have

$$
\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}\big(|X - \mathrm{E}(X)|^2\big) = \mathrm{E}\Big(\big(X - \mathrm{E}(X)\big)\big(X^* - \mathrm{E}(X^*)\big)\Big)\\
&= \mathrm{E}\big(XX^* - X\,\mathrm{E}(X^*) - X^*\,\mathrm{E}(X) + \mathrm{E}(X)\,\mathrm{E}(X^*)\big)\\
&= \mathrm{E}(XX^*) - \mathrm{E}(X)\,\mathrm{E}(X^*) - \mathrm{E}(X^*)\,\mathrm{E}(X) + \mathrm{E}(X)\,\mathrm{E}(X^*)\\
&= \mathrm{E}(|X|^2) - \mathrm{E}(X)\big(\mathrm{E}(X)\big)^* - \big(\mathrm{E}(X)\big)^*\,\mathrm{E}(X) + \mathrm{E}(X)\big(\mathrm{E}(X)\big)^* = \mathrm{E}(|X|^2) - \big|\mathrm{E}(X)\big|^2.
\end{aligned}
$$

in complete analogy with the situation when X is real valued.

16.4   The spectrum of a stationary process

Assume $\{Y_t\}$ is a real-valued stationary process, and write $\mu = \mathrm{E}(Y_t)$. Writing

$$Y_t = \sum_{k=0}^{N-1} C_k e^{2ikt\pi/N}. \qquad (0 \le t < N).$$

the analog of formula (16.2) becomes

(16.4) $$C_k = \frac{1}{N}\sum_{t=0}^{N-1} Y_t e^{-2ikt\pi/N} \qquad (0 \le k < N).$$

Taking expectations, we obtain

(16.5) $$\mathrm{E}(C_k) = \frac{1}{N}\sum_{t=0}^{N-1} E(Y_t) e^{-2ikt\pi/N} = \begin{cases} \mu & \text{if } k = 0, \\ 0 & \text{if } 1 \le k \le N-1 \end{cases}$$

according to (14.3). So we also have

$$C_k = \frac{1}{N}\sum_{t=0}^{N-1} Y_t e^{-2ikt\pi/N} = \frac{1}{N}\sum_{t=0}^{N-1}(Y_t - \mu) e^{-2ikt\pi/N} \qquad (1 \le k < N);$$

the second equation here holds according to (14.3). Hence, for k with $1 \le k < N$ we have

$$
\begin{aligned}
\mathrm{Var}(C_k) &= \mathrm{E}(|C_k|^2) - \big|\mathrm{E}(C_k)\big|^2 = \mathrm{E}(C_k^2) = \frac{1}{N^2}\,\mathrm{E}\left(\Big|\sum_{t=0}^{N-1}(Y_t - \mu)e^{-2ikt\pi/N}\Big|^2\right)\\
&= \frac{1}{N^2}\,\mathrm{E}\Big(\sum_{t=0}^{N-1}\sum_{t'=0}^{N-1}(Y_t - \mu)(Y_{t'} - \mu)e^{-2ik(t-t')\pi/N}\Big)\\
&= \frac{1}{N^2}\sum_{t=0}^{N-1}\sum_{t'=0}^{N-1} e^{-2ik(t-t')\pi/N}\,\mathrm{E}\big((Y_t - \mu)(Y_{t'} - \mu)\big)\\
&= \frac{1}{N^2}\sum_{t=0}^{N-1}\sum_{t'=0}^{N-1} e^{-2ik(t-t')\pi/N}\gamma(t - t').
\end{aligned}
$$

That is, putting $t = t' + m$, we have

$$N \operatorname{Var}(C_k) = \frac{1}{N} \sum_{m=-\infty}^{\infty} \gamma(m) e^{-2imk\pi/N} \sum_{t=\max(0,m)}^{\min(N-1,N-1+m)} 1.$$

Of course, the outside sum here is not a truly infinite sum, since the inside sum is empty (and therefore zero) for $|m| > N$. Writing $f = k/N$, we obtain

$$N \operatorname{Var}(C_{fN}) = \sum_{m=-N+1}^{N-1} \gamma(m) e^{-2ifm\pi} \frac{N - |m|}{N}.$$

Making $N \to \infty$, it follows that

$$(16.6) \qquad \lim_{N\to\infty} N \operatorname{Var}(C_{fN}) = \sum_{m=-\infty}^{\infty} \gamma(m) e^{-2ifm\pi},$$

assuming that this series $\sum_{m=-\infty}^{\infty} |\gamma(m)|$ is convergent. The limit on the left-hand side is called the power spectral density $S(f)$. Given that $\gamma(m) = \gamma(-m)$ is real, by (13.10) we have

$$(16.7) \qquad S(f) = \gamma(0) + 2 \sum_{m=1}^{\infty} \gamma(m) \cos(2fm\pi).$$

This result showing the existence of the spectrum is called the Wiener–Khinchin theorem.[16.5] The frequency function here defined is an even function with a period of 1; this means that its values in the interval $[0, 1/2)$ determine its values on the whole real line. Comparing these equations to equation (16.3), we expect that periodogram approximates twice the spectrum.[16.6] The problem of estimating the spectrum is, however, somewhat more complicated, as we will discuss below.

16.5 The periodogram is an inconsistent estimator of the spectrum

The question arises how good is the periodogram, described in equation (16.3), for estimating the spectrum. We will consider the special case of a white noise process. That is, let Y_t $(0 \le t < N)$ be independent normal $\mathcal{N}(0, \sigma^2)$ variables. Let $N > 3$ be an integer (we expect N to be fairly large), and $1 \le k < N/2$. For the sake of simplicity, assume N is odd.[16.7] Then C_k defined by equation (16.4) is a complex-valued random variable.

Writing

$$A_k = \frac{2}{N} \sum_{t=0}^{N-1} Y_t \cos \frac{2kt\pi}{N} \quad \text{and} \quad B_k = \frac{2}{N} \sum_{t=0}^{N-1} Y_t \sin \frac{2kt\pi}{N}$$

in analogy with equations (14.13). with $x_t = 2\pi t/N$ and $f(x_t) = Y_t$, these equations imply that the joint distribution of (A_k, B_k) is a multivariate normal distribution according to the definition

---

[16.5] What we stated is only the discrete case. The Wiener–Khinchin theorem is about a more general class of stochastic processes.

[16.6] We have $\operatorname{E}(|C_k|^2) = \operatorname{Var} C_k$, since $\operatorname{E}(C_k) = 0$ according to equation (16.5).

[16.7] All the arguments that follow will also work in the case of even N. The only reason that we assume N is odd is that we will refer to real interpolation formulas with N nodes, and the case of an odd number of nodes is somewhat simpler than the case of even number of nodes. In any case, when referring to formulas involving interpolation, we would have to refer to different formulas when the number of nodes is odd or even.

given in equation (2.4). These equations are connected with C_k with the equations $A_k = 2\Re C_k$ and $B_k = -2\Im C_k$ according to equations (14.12).

It is not hard to calculate the variances of A_k and B_k:

$$
\begin{aligned}
\mathrm{Var}(A_k) = \mathrm{E}\big((A_k)^2\big) &= \frac{4}{N^2}\,\mathrm{E}\bigg(\bigg(\sum_{t=0}^{N-1} Y_t \cos\frac{2kt\pi}{N}\bigg)^2\bigg) \\
&= \frac{4}{N^2}\,\mathrm{E}\bigg(\sum_{t=0}^{N-1}\sum_{t'=0}^{N-1} Y_t \cos\frac{2kt\pi}{N}\, Y_{t'}\cos\frac{2kt'\pi}{N}\bigg) \\
&= \frac{4}{N^2}\sum_{t=0}^{N-1}\sum_{t'=0}^{N-1}\mathrm{E}(Y_t Y_{t'})\cos\frac{2kt\pi}{N}\cos\frac{2kt'\pi}{N} \\
&= \frac{4}{N^2}\sum_{t=0}^{N-1}\mathrm{E}(Y_t^2)\cos^2\frac{2kt\pi}{N} = \frac{4}{N^2}\sum_{t=0}^{N-1}\sigma^2\cos^2\frac{2kt\pi}{N} = \frac{2}{N}\sigma^2;
\end{aligned}
$$

the fourth equation holds because Y_t and $Y_{t'}$ are independent for $t \neq t'$ and the last equation holds in view of the first equation in (14.15), and because $\mathrm{Var}(Y_t) = \sigma^2$. Similarly, we have

$$
\begin{aligned}
\mathrm{Var}(B_k) = \mathrm{E}\big((B_k)^2\big) &= \frac{4}{N^2}\,\mathrm{E}\bigg(\bigg(\sum_{t=0}^{N-1} Y_t \sin\frac{2kt\pi}{N}\bigg)^2\bigg) \\
&= \frac{4}{N^2}\,\mathrm{E}\bigg(\sum_{t=0}^{N-1}\sum_{t'=0}^{N-1} Y_t \sin\frac{2kt\pi}{N}\, Y_{t'}\sin\frac{2kt'\pi}{N}\bigg) \\
&= \frac{4}{N^2}\sum_{t=0}^{N-1}\sum_{t'=0}^{N-1}\mathrm{E}(Y_t Y_{t'})\sin\frac{2kt\pi}{N}\sin\frac{2kt'\pi}{N} \\
&= \frac{4}{N^2}\sum_{t=0}^{N-1}\mathrm{E}(Y_t^2)\sin^2\frac{2kt\pi}{N} = \frac{4}{N^2}\sum_{t=0}^{N-1}\sigma^2\sin^2\frac{2kt\pi}{N} = \frac{2}{N}\sigma^2;
\end{aligned}
$$

the fourth equation holds because Y_t and $Y_{t'}$ are independent for $t \neq t'$ and the last equation holds in view of the second equation in (14.15).

We next show that $\mathrm{Cov}(A_k, B_k) = 0$. We have

$$
\begin{aligned}
\mathrm{Cov}(A_k, B_k) = \mathrm{E}(A_k B_k) &= \frac{4}{N^2}\,\mathrm{E}\bigg(\sum_{t=0}^{N-1}\sum_{t'=0}^{N-1} Y_t \cos\frac{2kt\pi}{N} Y_{t'}\sin\frac{2kt'\pi}{N}\bigg) \\
&= \frac{4}{N^2}\sum_{t=0}^{N-1}\sum_{t'=0}^{N-1}\mathrm{E}(Y_t Y_{t'})\cos\frac{2kt\pi}{N}\sin\frac{2kt'\pi}{N} \\
&= \frac{4}{N^2}\sum_{t=0}^{N-1}\mathrm{E}(Y_t^2)\cos\frac{2kt\pi}{N}\sin\frac{2kt\pi}{N} = \frac{4}{N^2}\sum_{t=0}^{N-1}\sigma^2\cos\frac{2kt\pi}{N}\sin\frac{2kt\pi}{N} = 0;
\end{aligned}
$$

the fourth equation holds because Y_t and $Y_{t'}$ are independent for $t \neq t'$ and the last equation holds in view of the third equation in (14.15). Thus A_k and B_k, being uncorrelated random variables with a joint multivariate normal distribution, are independent in view of Theorem 2.1.

Thus, the the covariance matrix of (A_k, B_k) is

$$
\begin{pmatrix} 2\sigma^2/N & 0 \\ 0 & 2\sigma^2/N \end{pmatrix}.
$$

As we have $C_k = (A_k - iB_k)/2$ according to equation (14.12), we have $\mathrm{E}(N|C_k|^2) = \mathrm{E}\big(N(A_k^2 + B_k^2)/4\big) = \sigma^2$. Thus $N|C_k|^2 = N(A_k^2 + B_k^2)/4$ has a $\chi^2(2)$ distribution[16.8] with expectation σ^2. Thus $N|C_k|^2$ has the distribution of $\sigma^2/2$ times a standard χ^2 variable of degree of freedom 2. Hence its variance is $N|C_k|^2$ is $(\sigma^2/2)^2 \cdot 4 = \sigma^4$, since the variance of the standard χ^2 variable of degree of freedom 2 is 4 (see footnote 16.8 on p. 61). So, using the sample value c_k of C_k to estimate the spectrum gives poor results (cf. equations (16.3), (16.6), and (16.7)). In statistical language, the periodogram is an inconsistent estimator of the spectrum.[16.9]

16.6 Estimating the spectrum

As we we have seen $N|c_k|^2$ is an inconsistent estimator of the spectrum $S(f)$ with $f = k/N$. To develop a better estimator, we simply average neighboring values. That is, let $\lambda_{n,N}$ be a sequence of nonnegative of nonnegative numbers such that

$$\sum_{n=-\infty}^{\infty} \lambda_{n,N} = 1;$$

here, usually only a finite number of the $\lambda_{n,N}$ is nonzero; often, the best choice is to make about $\sqrt{N}$ of them to be nonzero. Then, instead of $|c_k|^2$, we use

$$\sum_{n=-\infty}^{\infty} \lambda_{k,N}|c_{k-n}|^2$$

to estimate the spectrum.[16.10] The sequence $\{\lambda_{n,N}\}_n$ is called a spectral window. Usually, the value of $\lambda_{n,N}$ is the largest for n close to zero. Most often, the window is symmetric, that is we have $\lambda{-n,N} = \lambda_{n,N}$. There are a great variety of spectral windows in use.

17 Orthogonal systems of functions

17.1 Inner product spaces

Let V be a vector space over F, where F is either the set of real numbers $\mathbb{R}$ or the set of complex numbers $\mathbb{C}$.

Definition 17.1. An inner product is a mapping $\langle \cdot, \cdot \rangle : V \times V \to F$ such that

(a) For all $x \in V$, $\langle x, x \rangle \geq 0$, and $\langle x, x \rangle = 0$ only if $x = 0$,

(b) $\langle x, y \rangle = \langle y, x \rangle^*$ for all $x, y \in V$,

(c) $\alpha\langle x, y \rangle = \langle x, \alpha y \rangle$ for all $\alpha \in F$ and $x, y \in V$,

(d) $\langle x, y \rangle + \langle x, z \rangle = \langle x, y + z \rangle$ for all $x, y, z \in V$.

[16.8] I.e., a χ^2 distribution with degree of freedom 2. The standard χ^2 distribution of degree of freedom k is defined as the sum the squares of k independent standard normal random variables. Its mean is k and its variance is $2k$. A constant multiple of such a variable is called a χ^2 variable of degree of freedom k.

[16.9] Given a parameter θ and an estimator $\hat{\theta}$ calculated from a sample, $\hat{\theta}$ is said to be a *consistent* estimator if $\hat{\theta}$ converges to θ in probability when the sample size goes to infinity. calculating the periodogram in formula (16.3), a sample size of N is used to calculate c_k.

[16.10]Note that this sum is the convolution of the sequences $\{\lambda_{n,N}\}$ and $\{|c_n|^2\}$.

A vector space with an inner product is called an *inner product space.*

In Clause (a), $\langle x, x\rangle \geq 0$ means that the complex number $\langle x, x\rangle$ is actually a nonnegative real. According to Clauses (b) and (c), we have $\alpha\langle x, y\rangle = \langle \alpha^* x, y\rangle$. If $F = \mathbb{R}$, the complex conjugation has no effect. In Section 13, we already dealt with several inner product spaces. When discussing Fourier series, the vector space was the set of complex-valued or real-valued functions f on the interval $[-\pi, \pi)$ for which $\int_{-\pi}^{\pi} |f|^2$ exists, and the inner product was

$$\langle f, g\rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} (f(x))^* \, g(x) \, dx.$$

For complex Fourier series, we used $F = \mathbb{C}$, for the real version we used $F = \mathbb{R}$. The factor $1/(2\pi)$ in front of the integral is not essential, but it makes the discussion more elegant in terms of inner products. For the real case, it is best to use the factor $1/\pi$ instead.[17.1] When discussing trigonometric interpolation, the inner product was vector space was the set of complex-valued or real-valued functions on the set $\{x_k : 0 \leq k < N\}$, and the inner product was

$$\langle f, g\rangle = \frac{1}{N} \sum_{n=0}^{N-1} (f(x_n))^* \, g(x_n).$$

Another example for an inner-product space is the set of complex-valued random variables on a given probability space. For random variables X and Y, we take $\langle X, Y\rangle = \mathrm{E}(X^*Y)$. In order to make sure that Clause (a) is satisfied, the random variables X and Y must be considered equal if $\mathrm{P}(X = Y) = 1$.[17.2] Similarly, the real-valued random variables give rise to an inner product space over the reals if we take $\langle X, Y\rangle = \mathrm{E}(XY)$.

Schwarz's inequality says that we have

(17.1) $$|\langle x, y\rangle|^2 \leq \langle x, x\rangle \cdot \langle y, y\rangle;$$

see Problem 17.1 below. On a vector space V over F (with $F = \mathbb{C}$ or $\mathbb{R}$) one often defines a norm:

Definition 17.2. A *norm* is a mapping $\|\cdot\| : V \to \mathbb{R}$ such that

(a) $\|x\| \geq 0$ for all $x \in V$, and $\|x\| = 0$ only if $x = 0$,

(b) $\|\alpha x\| = |\alpha| \, \|x\|$ for all $\alpha \in F$ and for all $x \in V$,

(c) $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in V$.

A vector space with a norm is called a *normed vector space* or, more shortly, a *normed space.*

Clause (c) is called Minkowski's inequality. With an inner product $\langle \cdot, \cdot\rangle$ one can define the induced norm as $\|x\| = \sqrt{\langle x, x\rangle}$. If the norm is induced by an inner product, Minkowski's inequality can be proved by Schwarz's inequality; see Problem 17.2 below. In an inner product space, by the norm we will always mean the induced norm unless otherwise mentioned.

[17.1] For a proper discussion of these function spaces (i.e., the vector spaces just described), Riemann integration is not really an adequate tool, and one needs to use the newer integral concept invented by Henri Lebesgue in 1904. However, we will not get into subtle issues of convergence where the advantages of Lebesgue integration are felt.

[17.2] That is, the inner-product space is formed by the equivalence classes of random variables under the equivalence relation $X \equiv Y$ if $P(X = Y) = 1$. Note that this ensures that Clause (a) is satisfied; cf. Problem 5.1 above.

Two vectors x and y are called *orthogonal* if $\langle x, y \rangle = 0$. Let x_1, x_2, ..., x_n be a system of vectors such that x_i and x_j are orthogonal whenever $1 \leq i < j \leq n$. Then

$$(17.2) \qquad \left\| \sum_{k=1}^{n} x_k \right\|^2 = \sum_{k=1}^{n} \|x_k\|^2.$$

Indeed, we have

$$\left\| \sum_{k=1}^{n} x_k \right\|^2 = \left\langle \sum_{k=1}^{n} x_k, \sum_{l=1}^{n} x_l \right\rangle = \sum_{k=1}^{n} \sum_{l=1}^{n} \langle x_k, x_l \rangle = \sum_{k=1}^{n} \|x_k\|^2;$$

the last equation holds since $\langle x_k, x_l \rangle = 0$ unless $k = l$. The equation we just established can be considered an analog of the Pythagorean theorem.

17.2   Orthonormal systems

Definition 17.3. A system of vectors $S = \{f_1, f_2, f_3, \ldots\}$ is called *orthonormal* if

$$(17.3) \qquad \langle f_k, f_l \rangle = \delta_{kl}.$$

It is called *complete* if every vector $f \in V$ can be expressed as

$$(17.4) \qquad f = \sum_k \alpha_k f_k.$$

If the orthonormality condition (17.3) is weakened to say that $\langle f_k, f_k \rangle > 0$ and $\langle f_k, f_l \rangle = 0$ if $k \neq l$ then the system is called orthogonal rather than orthonormal. The system S in this definition may be finite or infinite. In case of Fourier series, we had an infinite orthonormal system (the factor $1/(2\pi)$ in front of the integral above in the complex case and $1/\pi$ in the real case was needed to make the system orthonormal rather than only orthogonal), and in case of trigonometric interpolation we had a finite orthonormal system. If S is infinite, we need a concept of convergence to interpret the sum (17.3). In an inner product space there are several notions of convergence; the simplest we can use in this case is convergence in norm:

Definition 17.4. Assume V is a vector space with norm $\| \cdot \|$. Let $f \in V$ and let $\{f_n\}_{n=1}^{\infty}$ be a sequence, where $f_n \in V$. Then we say that f_n converges to f in norm if

$$\lim_{n \to \infty} \|f_n - f\| = 0.$$

We say

$$\sum_{n=1}^{\infty} f_n = f$$

if the partial sums

$$\sum_{k=1}^{n} f_k$$

converge to f in norm.

A sequence $f \in V$ and let $\{f_n\}_{n=1}^{\infty}$, where $f_n \in V$, called a *Cauchy sequence* if

$$\lim_{\substack{m \to \infty \\ n \to \infty}} \|f_m - f_n\| = 0.$$

A normed vector space is called *complete* if every Cauchy sequence is convergent.

The trigonometric system (in the complex and real Fourier series) is complete. Similarly, the corresponding finite systems in the cases of trigonometric interpolation considered above are complete. The inner product spaces considered in these examples are all complete with respect to the norm induced by the inner product in question.[17.3] In equation (17.4), the orthonormality relations imply that

$$(17.5) \qquad \alpha_k = \langle f_k, f \rangle.$$

This equation is the general statement of equations (13.4), (13.15), (14.6), and (14.13). We have the following

Lemma 17.1. *Let V be an inner product space, $S = \{f_1, f_2, f_3, \ldots\}$ be an orthonormal system of vectors in V and let $f \in V$ be an arbitrary vector, and let $\alpha_k = \langle f_k, f \rangle$. Then*

$$(17.6) \qquad \sum_k |\alpha_k|^2 \leq \|f\|^2.$$

We have equality here if and only if

$$(17.7) \qquad f = \sum_k \alpha_k f_k.$$

Inequality (17.6) is called Bessel's inequality. When we have equality in Bessel's inequality, we obtain Parseval's identity (more on this below):

$$(17.8) \qquad \|f\|^2 = \sum_k |\alpha_k|^2.$$

This is the general statement of the Parseval identities (13.16), (13.17), (14.6), and (14.17).

Proof. Write $S = \{f_k : k < m\}$, where m is an integer or $m = \infty$. For any integer $n \leq m$ we write

$$g_n = \sum_{k=1}^{n} \alpha_k f_k.$$

Then we have $\langle f_k, g_n \rangle = \alpha_k$ for any $k \leq n$, and so

$$\langle f_k, f - g_n \rangle = \langle f_k, f \rangle - \langle f_k, g_n \rangle = \alpha_k - \alpha_k = 0$$

for every $k \leq n$. Hence any two of the vectors $f - g_n$ and $\alpha_k f_k$ ($k \leq n$) are orthogonal. Therefore according to equation (17.2) we have

$$(17.9) \qquad \|f\|^2 = \|f - g_n\|^2 + \sum_{k=1}^{n} \|\alpha_k f_k\|^2 = \|f - g_n\|^2 + \sum_{k=1}^{n} |\alpha_k|^2.$$

This establishes inequality (17.6).

According to formula (17.9), equality in (17.6) means that $\|f - g_n\| = 0$ for $n = m$ if m is finite, or that

$$\lim_{n \to \infty} \|f - g_n\| = 0$$

if m is infinite. In either case, this is equivalent to saying that (17.7) holds. $\qquad \square$

[17.3]For the space associated with Fourier series, see the discussion below, on p. 65, especially footnote 17.4 on the same page.

Corollary 17.1. *Let V be an inner product space, $S = \{f_1, f_2, f_3, \ldots\}$ be an orthonormal system of vectors in V. Assume there is no nonzero vector $g \in V$ such that $\langle f_k, g \rangle = 0$ for all f_k in S. If S is finite, or if S is infinite and V is complete space, then S is complete orthonormal system.*

Note that in this corollary we have the extra assumption about the completeness of V. This was done in order to guarantee the convergence of the series

$$\sum_k \alpha_k f_k \qquad (\alpha_k = \langle f_k, f \rangle),$$

in case S is infinite. In (17.7), the convergence of the series on the right-hand side was guaranteed, since we assumed that equality holds there. Without the assumption of equality, the conververgence of this series is not guaranteed. However, if we assume that V is complete, the the convergence of this series follows. Indeed, assume that S is infinite. Inequality (17.6) implies that $\sum_{k=1}^{\infty} |\alpha_k|^2$ is convergent. Writing

$$g_n = \sum_{k=1}^{n} \alpha_k f_k.$$

for the partial sums of the series $\sum_{k=1}^{\infty} \alpha_k f_k$, given integers μ and ν with $1 \leq \mu < \nu$, we have

$$\|g_\nu - g_\mu\|^2 = \Big\| \sum_{k=\mu+1}^{\nu} \alpha_k f_k \Big\|^2 = \sum_{k=\mu+1}^{\nu} \|\alpha_k f_k\|^2 = \sum_{k=\mu+1}^{\nu} |\alpha_k|^2 \|f_k\|^2 = \sum_{k=\mu+1}^{\nu} |\alpha_k|^2;$$

the second equality here holds according to (17.2). This shows that $\sum_{k=1}^{\infty} \alpha_k f_k$ is a Cauchy sequence; hence it is convergent, since we assumed that V is complete.

Proof. Assume S is not a complete orthonormal system. Then, according to the assumptions, there is an $f \in V$ such that equation (17.7) does not hold for this f, i.e., that

(17.10)
$$g \stackrel{def}{=} f - \sum_k \langle f_k, f \rangle f_k \neq 0.$$

Now, for any f_l in S we have

$$\langle f_l, g \rangle = \langle f_l, f \rangle - \sum_k \langle f_l, \langle f_k, f \rangle f_k \rangle = \langle f_l, f \rangle - \sum_k \langle f_k, f \rangle \langle f_l, f_k \rangle$$

$$= \langle f_l, f \rangle - \sum_k \langle f_l, f \rangle \delta_{lk} = \langle f_l, f \rangle - \langle f_l, f \rangle = 0;$$

to rigorously establish the second equation here, some convergence issues need to be dealt with, but these are easily handled with Schwarz's inequality (17.1) – see Problem 17.3 below. This is a contradiction, since we assumed that no vector g exists for which $\langle f_l, g \rangle = 0$ for all f_l in S. □

The space of functions that are square integrable on the interval (a, b), called $L^2(a, b)$, space is an important example of a complete space.[17.4] Fourier series were considered on the space $L^2(-\pi, \pi)$.

[17.4]The name $L^2(a, b)$ does not specify whether the functions are real valued or complex valued. When it is not clear from the context, one can make the distinction by calling it a real L^2 space or a complex L^2 space. The completeness of these spaces (for any interval (a, b), finite or infinite) is the Riesz-Fischer theorem. It was proved independently by Frigyes (Frederick) Riesz and Ernst Sigismund Fischer. The integral concept used in this theorem is that of Lebesgue – the result is not true with Riemann integration.

In actual fact, the elements of $L^2(a, b)$ are not functions; they are equivalence classes for functions under the

17.2.1 Gram–Schmidt orthogonalization

Let V be a vector space, let m be a positive integer and let $S = \{f_k : 1 \le k < m\}$ be an orthonormal system of vectors in V. Assume S is not complete. Then, as we saw in the proof of Corollary 17.1 there is a vector $f \in V$ such that the inequality in formula (17.10) holds. Taking $f_m = (1/\|g\|)g$ with the g defined in this formula, the system $S' = S \cup \{f_m\}$ is orthonormal. If V is finite dimensional, then starting with $S = \emptyset$, we can obtain a complete orthonormal system in finitely many steps.

If V is infinite dimensional, then, in order to obtain a complete orthonormal system one needs to proceed more delicately, because even after repeating this step infinitely many times, the resulting infinite system may not be complete. While this issue is only of marginal interest for our purposes, we will outline one possible way we may proceed in this case. For this, we need the following

Definition 17.5. Let V be an inner product space and let $M \subset V$. We say that M is dense in V if for every $\epsilon > 0$ and for every $f \in V$ there is a $g \in M$ such that $\|f - g\| < \epsilon$.

We recall that given a vector space V and a subset D, the span of D is the smallest subspace of V including D. It is well known that the elements of the span of D are exactly the *finite* linear combinations of the elements of D.[17.5]

Lemma 17.2. *Let V be an inner product space, $S = \{f_1, f_2, f_3, \ldots\}$ be an orthonormal system of vectors in V and let $f \in V$. Let M be a set such that the span of M is dense in V, and assume that equation (17.4) holds for every $f \in M$. Then S is complete.*

Proof. By linearity, equation (17.4) holds for all f in the span of M, and then, by taking limits, we can conclude that this equation holds for every $f \in V$. Hence S is complete. $\qquad\square$

If $M = \{h_k : 1 \le k < \infty\}$ is a subset of V such that the span of M is dense in V, then one can modify the above method to obtain an orthonormal system in V as follows. Let $m > 0$ be an integer and assume the orthonormal system $S_m = \{f_k : 0 \le k < m\}$ has already been constructed. Pick the least positive integer l such that $S_m \cup \{h_l\}$ is linearly independent.[17.6] If such an l can be found, then writing,

$$(17.11) \qquad\qquad g_m \stackrel{def}{=} h_l - \sum_{k=1}^{m-1} \langle f_k, h_l \rangle f_k \neq 0.$$

put $S_{m+1} = S_m \cup \{f_m\}$ with $f_m = (1/\|g_m\|)g_m$. If no such l can be found, put $S_{m+1} = S_m$. Then the system

$$S = \bigcup_{m=1}^{\infty} S_m$$

is a complete orthonormal system. The reason is that the construction ensures that equation (17.4) holds for every $f \in M$; hence the completeness of S follows from Lemma 17.2.

---

equivalence relation

$$f \sim g \equiv \int_a^b |f - g|^2 = 0.$$

It is necessary to take equivalence classes in order to make sure that Clause (a) of Definition 17.1 is satisfied. It is common parlance, however, to talk about elements for $L^2(a, b)$ as functions rather than equivalence classes of functions. Functions that belong to the same equivalence class are said to be equal a.e. (almost everywhere).

[17.5]Linear combination always means finite linear combinations, unless otherwise indicated. We included the word "finite" for emphasis, since we have considered infinite sums above.

[17.6]It is easy to see that any orthonormal system is linearly independent, and so S_m itself is linearly independent.

There are several well-known countable dense subsets of $L^2(a,b)$. The simplest one is formed by the finite linear combinations with rational coefficients of the characteristic functions of all finite open intervals with rational endpoints.[17.7]

17.3   Problems

Problem 17.1. Given two vectors in an inner product space V over $\mathbb{R}$ or $\mathbb{C}$, show that

$$\left|\langle x, y\rangle\right|^2 \le \langle x, x\rangle\langle y, y\rangle$$

for any $x, y \in V$. (This inequality is called Schwarz inequality).

Problem 17.2. Let V be an inner product space over $\mathbb{R}$ or $\mathbb{C}$, and for $x \in V$ define its norm as $\|x\| = \sqrt{\langle x, x\rangle}$. Show that for any $x, y \in V$ we have

$$\|x + y\| \le \|x\| + \|y\|.$$

(This inequality is called Minkowski's inequality.)

Problem 17.3. Let V be an inner product space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\| \cdot \|$, and let f and f_n for all $n > 0$ be elements of V. Assume that

$$\lim_{n \to \infty} \|f_n - f\| = 0.$$

Show that for all $g \in V$ we have

$$\lim_{n \to \infty} \langle g, f_n - f\rangle = 0.$$

Hint: Use Schwarz's inequality.

Problem 17.4. Let V be a vector space over $\mathbb{C}$, and let $\langle \cdot, \cdot \rangle_{\mathbb{R}}$ be a real-valued inner product on V considered as a vector space over $\mathbb{R}$ (that is, Clause (c) in Definition 17.1 is only assumed for real α) with the additional property that

(17.12) $$\langle if, ig\rangle_{\mathbb{R}} = \langle f, g\rangle_{\mathbb{R}} \qquad \text{for all} \quad f, g \in V.$$

Show that

(17.13) $$\langle f, g\rangle = \langle f, g\rangle_{\mathbb{R}} + i\langle if, g\rangle_{\mathbb{R}}$$

is a complex inner product on V over $\mathbb{C}$.

Problem 17.5. Let V be a normed vector space over $\mathbb{R}$ or $\mathbb{C}$. Show that

(17.14) $$\left|\|f\| - \|g\|\right| \le \|f - g\| \qquad \text{for all} \quad f, g \in V.$$

Problem 17.6. Let V be a normed vector space over $\mathbb{R}$ or $\mathbb{C}$, and let $f \in V$ and $f_n \in V$ for all positive integers n. If $f_n \to f$ in norm, show that

(17.15) $$\lim_{n \to \infty} \|f_n\| = \|f\|.$$

[17.7]The characteristic function of a set U is a function that is 1 in U and zero elsewhere. We need to take rational endpoints to make sure that the set of functions we obtain is countable.

18 Building ARIMA models directly

Let $\{Y_t\}$ be a time series, and assume observed values y_n are available for times $1 \leq n \leq N$, where $N > 0$ is an integer; we assume that Y_t is real, but it is easy to extend these considerations to complex-valued time series. Let p, q, and d be integers such that $0 \leq d < p$ and $q \geq 0$. We would like to build an ARIMA$(p-d, d, q)$ model for Y_t. We expect that d is large enough such that $\{\nabla^d Y_t\}$ is stationary with zero means, but we do not wish to determine d. Determine the AR coefficients in equation

$$(18.1) \qquad Y_t = \sum_{k=1}^{p} \phi_k Y_{t-k} + e_t - \sum_{k=1}^{q} \theta_k e_{t-k}$$

by using least square approximation; i.e., let ϕ_k for k with $1 \leq k \leq p$ be such that

$$(18.2) \qquad \sum_{n=p+1}^{N} \frac{1}{1 + y_n^2} \left(y_n - \sum_{k=1}^{p} \phi_k y_{n-k} \right)^2$$

be the least possible.[18.1] Write

$$(18.3) \qquad \phi(x) = 1 - \sum_{k=1}^{p} \phi_k x^k.$$

With the notation introduced in equation (18.3), this suggests the approximate AR model

$$(18.4) \qquad \phi(B)Y_t \approx 0.$$

We wrote $\approx$ instead of $=$, since p is not large enough to build a good ARI model (where as p and q together should be suitable to build an ARIMA model). What is meant by $\approx$ here is unimportant, since this equation will not be used, it will only be a guide as to how to build the ARIMA model.

To determine the MA coefficients, let $m \leq N-p-q$ be a large positive integer. We will comment on the choice of m later. Put

$$(18.5) \qquad \mathbf{y}_k = (y_{k-m+1}, y_{k-m+2}, \ldots, y_k)^T \qquad (k \geq m).$$

We determine the error vectors $\mathbf{e}_t$ by orthogonalizing the vectors $\mathbf{y}_t$ for t with $m \leq t \leq N$ with respect to the real inner product $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}$ *without normalizing*. That is, we put $\mathbf{e}_m = \mathbf{y}_m$ and assuming that $\mathbf{e}_k$ has been defined for k with $m \leq k < t$, where $m < t \leq N$, we put

$$(18.6) \qquad \mathbf{e}_t = \mathbf{y}_t - \sum_{\substack{k=1 \\ \mathbf{e}_k \neq 0}}^{t-1} \frac{1}{\mathbf{e}_k^T \mathbf{e}_k} (\mathbf{e}_k^T \mathbf{y}_k) \mathbf{e}_k,$$

where we sum only for those values of k for which $\mathbf{e}_k \neq 0$ (in which case $\mathbf{e}_k^T \mathbf{e}_k \neq 0$, so we do not have a zero in the denominator).[18.2] As for the choice of m, there is a danger in choosing N too small as compared to N, since there are about $N - m$ vectors $\mathbf{y}_t$, and this number needs to be

---

[18.1]Since we do not assume that the time series $\{Y_t\}$ is stationary, we need to allow larger errors if the value of y_n is large. We divide by $1 + y_n^2$ instead of y_n^2 to avoid dividing by zero in case y_n is zero.

[18.2]The parentheses on the right-hand side of equation (18.6) were only written for clarity; they are not needed, since matrix multiplication is associative.

substantially smaller than the length m of the vectors so that the orthogonalized vectors retain a random character. There are m linearly independent vectors of length m, so after orthogonalizing m linearly independent vectors, the whole space of m-dimensional vectors will be spanned. Perhaps the choice of $m \approx N - \sqrt{N}$ is the optimal. This makes the number of vectors to be orthogonalized to be about $\sqrt{m}$.

Equation (18.6) should be compared to (17.11) of Gram–Schmidt orthogonalization; the difference is that in that equation we have $\|f_k\| = 1$, whereas here we do not require that $\|e_k\| = 1$, where the norm is the norm induced by the inner product.[18.3] This kind of orthogonalization without normalizing was described earlier on account of the innovations algorithm; see Section 7.

Equation (18.6) is not to be used in numerical calculations. This is because, we mentioned at the end of Subsection 7.1, the Gram–Schmidt orthogonalization is numerically unstable; that is, small numerical errors committed initially give rise to large errors later in the calculation. There are also numerical problems with doing least square optimization in the common sense way, that is, by taking the partial derivatives of the expression describing the least squares error, and looking for its minimum by equating the partial derivatives to zero. Interestingly, both of these two problems can be handled in a stable way by the QR decomposition using Householder transformations. The QR decomposition starts with an $m \times n$ matrix A, where $m \geq n > 0$ and finds an orthogonal matrix[18.4] such that the equation

$$QA = R', \qquad \text{where} \quad R' = \begin{pmatrix} R \\ 0_{m-n,n} \end{pmatrix}$$

holds, where R is an upper triangular matrix, and $0_{m-n,n}$ is the $(m-n) \times n$ zero matrix. In [23, Section 38, pp. 174–184] it is explained how the QR decomposition can be used to solve the least squares optimization problem. As for using it to solve the orthogonalization problem, the orthonormal vectors resulting from the orthogonalization of the columns of the matrix A will be the rows of the matrix Q, i.e., the columns of the matrix $Q^{-1} = Q^T$.[18.5] The coefficients to express the kth column of A as a linear combination the first k columns of the matrix Q^T are contained in the kth column of R' (or R, since all the coefficients in R' outside R are 0 – since R is upper triangular, only the first k entries in the kth column are nonzero); this is because we have $A = Q^{-1}R' = Q^T R'$. In [23], the QR decomposition is described for real matrices, but it is easy to adapt it for complex matrices.

The numerical method we described will produce a system of the orthonormal vectors $\mathbf{e}'_t = (1/\|e_t\|)\mathbf{e}_t$, and then we can recover the vectors $\mathbf{e}_t$ from these and the related coefficients.

The equations in (18.6) can be rearranged to express $\mathbf{y}_t$ as a linear combination $\mathbf{e}_t$ as follows to obtain

$$(18.7) \qquad \mathbf{y}_t = \sum_{l=0}^{t-m} \psi_{l,t}\mathbf{e}_{t-l} = \mathbf{e}_t + \sum_{l=1}^{t-m} \psi_{l,t}\mathbf{e}_{t-l}.$$

by orthonogonality. The equation $\psi_{0,t} = 1$ easily follows from equation (18.6). As for the other coefficients, writing $\|\mathbf{x}\| = (\mathbf{x}^T\mathbf{x})^{1/2} = \langle \mathbf{x}, \mathbf{x} \rangle$ for the norm induced by the inner product we are using, given t and l with $m \leq t \leq N$ and $0 \leq l \leq t - m$, we have

$$\psi_{l,t} = \mathbf{e}_{t-l}^T \mathbf{y}_t / \|\mathbf{e}_t\|$$

if $\mathbf{e}_t \neq 0$ (in which case case $\|\mathbf{e}_t\| \neq 0$); $\mathbf{e}_t = 0$ then we can define $\psi_{l,t}$ arbitrarily, except that we want to put $\psi_{0,t} = 1$ also in this case; these equations follow easily from the considerations in Section 17; see e.g. equations (17.3), (17.4), and (17.5).

[18.3]That is, $\|\mathbf{x}\|^2 = \mathbf{x}^T\mathbf{x} = \langle \mathbf{x}, \mathbf{x} \rangle$.
[18.4]Orthogonal matrices were defined before equation (8.7).
[18.5]The latter equation holds since Q is orthogonal.

Writing

$$(18.8) \qquad \psi_t(x) = \sum_{l=0}^{t-m} \psi_{l,t} x^l,$$

the last equation can be written as

$$(18.9) \qquad \mathbf{y}_t = \psi_t(B)\mathbf{e}_t.$$

Multiplying this equation by the polynomial $\phi(B)$, where $\phi(x)$ is given in equation (18.3), we obtain

$$(18.10) \qquad \phi(B)\mathbf{y}_t = \phi(B)\psi_t(B)\mathbf{e}_t.$$

It is important to point out that the operator B in $\phi(B)$ acts on everything to the right of it, even on t in the subscript of the polynomial $\psi_t(B)$. That is, we have $B\psi_t(B)e_t = \phi_{t-1}(B)e_{t-1}$.

To explain the scope of B in $\phi(B)$, we need to write out how to obtain equation (18.10) in more detail. To simplify the notation, writing $\mathbf{y}_t = \mathbf{e}_t = 0$ for $t < m$, in equation (18.7) we can extend the summation to infinity:

$$\mathbf{y}_t = \sum_{l=0}^{\infty} \psi_{l,t}\mathbf{e}_{t-l}.$$

Note that this equation is identical to equation (18.9). Substituting this (with t or $t-k$ replacing t) into the expression

$$\phi(B) = \mathbf{y}_t = \mathbf{y}_t - \sum_{k=1}^{p} \phi_k \mathbf{y}_{t-k}$$

(cf. equation (18.4)), we obtain the equation

$$\mathbf{y}_t - \sum_{k=0}^{p} \phi_k \mathbf{y}_{t-k} = \sum_{l=0}^{\infty} \psi_{l,t}\mathbf{e}_{t-l} - \sum_{k=1}^{p}\phi_k \sum_{l=0}^{\infty} \psi_{l,t-k}\mathbf{e}_{t-k-l}.$$

A shorter way to write this equation was given in equation (18.10) with the scope of B in $\phi(B)$ as described after that equation.

Changing the from vectors to random variables, this suggests the equation

$$(18.11) \qquad \phi(B)Y_t = \phi(B)\psi_t(B)e_t,$$

where e_t is the random variable describing the error committed by the process at time t. This would give an ARIMA model except for the dependence of ψ_t on t.

If the time series $\{Y_t\}$ can be modeled by an ARIMA model, we expect that the time series $\{\phi(B)Y_t\}$ is stationary, because in this case the polynomial $\phi(x)$ is expected to include a factor $(x-1)^d$ for which already the time series $\{(B-I)^d Y_t\}$ is stationary. Then, assuming that equation (18.11) correctly models the time series $\{Y_t\}$, the coefficients of $\theta(x)\psi_t(x)$ cannot depend on t; this is because the innovations algorithm described in Section 7 allows us to determine these coefficients from the moments $(\theta(B)Y_t)(\theta(B)Y_{t-k})$, where the score of the operator first operator B stops at the enclosing parenthesis (that is, it does not affect the Y_{t-k}), and these moments only depend on k, and not on t. Hence, if we choose N and m large enough, for appropriate values of p and q, the polynomials $\phi(B)$ (which depends on N) and $\phi(B)\psi_t(B)$ approximate polynomials[18.6] $\phi(B)$ and $\theta'(B)$ such that the equation

$$(18.12) \qquad \phi(B)Y_t = \theta'(B)e_t,$$

[18.6]Note that we cannot write $\phi(x)\psi_t(x)$ instead of $\phi(B)\psi_t(B)$, since, as we indicated above, B in $\phi(B)$ acts also on the subscript of $\psi_t(B)$, so the coefficients of $\phi(x)\psi_t(x)$ and $\phi(B)\psi_t(B)$ are not the same.

correctly models the time series $\{Y_t\}$. If in the polynomial $\theta'(B)$ we discard the terms of degree higher than q to obtain the polynomial $\theta(B)$,[18.7] then we obtain the ARIMA$(p-d,d,q)$ model

$$(18.13) \qquad \phi(B)Y_t = \theta(B)e_t.$$

Here d is the largest integer for which $(x-1)^d$ is a factor of $\phi(x)$.

The ARIMA model obtained this way will probably not be identical to the ARIMA model obtained by differentiation, and numerical experiments are needed to evaluate the quality of models obtained this way as opposed to ARIMA models obtained by differencing. The difference can be described as follows: The present model makes no direct assumption about the size of the errors, since the errors are determined according to the innovation algorithms, and not by fitting an ARMA model. When fitting an ARMA model, the least squares method assumes that these errors will be about equal size; a similar assumption is made if the maximum likelihood method is used, since the maximum likelihood method is also based on some kind of least square optimization according to formula (10.5). In an ARIMA model, the errors e_t do not assume differencing, since in equation ARMA: ARIMA eq the polynomial $\theta(x)$ is expected to have all its zeros outside the unit circle. That is, the expectation is that even in an ARIMA model the errors are of about the same size, even when the size of Y_t may increase rapidly. On the other hand, the norming factor $1/(1+y_n^2)$ used in the least squares optimization makes one to expect that the errors are proportional to y_n (at least for large values of y_n; this appears to be a much more reasonable expectation.

18.1 Adding a drift term and ensuring zero means of innovations

An ARIMA model with a drift term has the form

$$(18.14) \qquad Y_t = \delta + \sum_{k=1}^{p} \phi_k Y_{t-k} + e_t - \sum_{k=1}^{q} \theta_k e_{t-k}.$$

To determine the AR coefficients in such a model by using least square approximation find the values of $\hat{\delta}$ and of the coefficients ϕ_k for $1 \le k \le p$ such that

$$(18.15) \qquad \sum_{n=p+1}^{N} \frac{1}{1+y_n^2} \left(y_n - \hat{\delta} - \sum_{k=1}^{p} \phi_k y_{n-k} \right)^2;$$

see footnote 18.1 on p. 68 for an explanation of the reason to divide by $1+y^2$. With the notation introduced in equation (18.3), this suggests the approximate AR model

$$(18.16) \qquad \phi(B)Y_t \approx \hat{\delta}.$$

Similarly to equation (18.4), this equation will not play a direct role; it will only give an indication as to how to build the ARIMA model. We wrote approximate equality since the value of p is not large enough to build a correct ARI model.

To determine the MA coefficients, let $m \le N-p-q$ be a large positive integer as before,[18.8] and let $\mathbf{y}_k$ as given in equation (18.5), and let $\mathbf{u} = (1,1,\dots,1)^T$ be the m-dimensional column vector with all its entries 1. We determine the error vectors $\mathbf{e}_t$ by orthogonalizing the vectors $\mathbf{u}$ and $\mathbf{y}_t$ for t with $m \le t \le N$ (in this order, $\mathbf{u}$ being the first one) with respect to the real inner product

---

[18.7] In fact, if an ARIMA$(p-d,d,q)$ model is appropriate, then the coefficients of the terms of degree higher than q of $\theta'(B)$ should be near 0.

[18.8] The choice of m is similar to the choice before. That is, perhaps the the best choice is $m \approx N - \sqrt{N}$. See the discussion after equation (18.6).

$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}$ *without normalizing.* As before, with the aid of these vectors we can express the vector $\mathbf{y}_t$ as

$$(18.17) \qquad \mathbf{y}_t = \delta_t \mathbf{u} + \sum_{l=0}^{t-m} \psi_{l,t} \mathbf{e}_{t-l} = \delta_t \mathbf{u} + \mathbf{e}_t + \sum_{l=1}^{t-m} \psi_{l,t} \mathbf{e}_{t-l}.$$

With the notation introduced in equations (18.8), this can be written as

$$(18.18) \qquad \mathbf{y}_t = \delta_t + \psi_t(B) \mathbf{e}_t;$$

of course, the polynomial $\psi_t(x)$ now is different from what it was above, but we use the same notation. Multiplying this equation by the polynomial $\phi(B)$, where has the same look as $\phi(x)$ given in equation (18.3), but now it is the polynomial used in equation (18.16), we obtain

$$(18.19) \qquad \phi(B)\mathbf{y}_t = \phi(B)\delta_t \mathbf{u} + \phi(B)\psi_t(B)\mathbf{e}_t.$$

Here, in the first term on the right-hand side, the operator B in $\phi(B)$ acts on δ_t, but not on $\mathbf{u}$, since the latter does not depend on t. Similarly as we explained after equation (18.10), in the second term the scope of B in $\phi(B)$ to the right of it.

Similarly as above, changing the from vectors to random variables, this suggests the equation

$$(18.20) \qquad \phi(B)Y_t = \phi(B)\delta_t + \phi(B)\psi_t(B)e_t,$$

where e_t is the random variable describing the error committed by the process at time t. This would give an ARIMA model with drift except for the dependence of ψ_t on t. As we explained above on account of the model without a drift term, if the time series $\{\phi(B)Y_t\}$ is stationary, then the polynomials in this equation should not depend on t. So, choosing N and m large enough, we will approximate a model

$$(18.21) \qquad \phi(B)Y_t = \phi(B)\delta_t + \theta'(B)e_t,$$

as in equation (18.12). Here δ_t may depend on t, but $\phi(B)\delta_t$ should not, as we will explain below. Writing δ for δ for $\phi(B)\delta_t$, and truncating $\theta'(B)$ by discarding the terms of degree higher than q, we obtain the ARIMA$(p-d, d, q)$ model

$$(18.22) \qquad \phi(B)Y_t = \delta + \theta(B)e_t.$$

Here d is the largest integer for which $(x-1)^d$ is a factor of $\phi(x)$. The constant δ on the right-hand side is called drift.

The difference between an ARIMA model with and without a drift term can be explained as follows. If we want to model the time series $\{Y_t\}$ with an ARIMA model via first building an ARMA model, we need to perform differencing on $\{Y_t\}$ until we obtain a stationary time series with zero means, and then build an ARMA model. If we include a drift term, then we do the differencing up to the point when we obtain a stationary time series $\{X_t\}$ but without requiring that $E(X_t) = 0$. Instead, we build an ARMA model for the time series $\{X_t - \mu\}$, where μ is an estimate for $E(X_t)$. If $X_t = (I - B)^d Y_t$ and $X_t - \mu$ is modeled as

$$X_t - \mu = \sum_{k=1}^{p-d} \tilde{\phi}_k X_{t-k} + \theta(B)e_t,$$

then the model in equation (18.13) can be written as

$$(I - B)^d Y_t = \mu\Big(1 - \sum_{k=1}^{p-d} \tilde{\phi}_k\Big) + \sum_{k=1}^{p-d} \tilde{\phi}_k B^k (I - B)^d Y_t + \theta(B)e_t.$$

The first term on the right-hand side is the drift term.

18.2 Seasonal ARIMA models

We will consider a multiplicative $\mathrm{ARIMA}(p, d, q) \times (P, D, Q)$ model with seasonal parameter s, which means that we have s equally timed observations per period. The form suggested for such a model in [4, Subsection 9.1,3, formula (9.1.7) on p. 332] is

$$\phi(B)\Phi(B^s)(I - B)^d(I - B^s)^D Y_t = \theta(B)\Theta(B^s)e_t;$$

see also [12, Section 10.2, p. 231]. Here the time series $\{(I - B)^d(I - B^s)^D Y_t\}$ is assumed to be stationary.

  We will describe how to build such a model. As in building the model in equation (18.9), we do not need to separate out the integration degrees d and D in advance. That is, given p, q, P, and Q, we will build an $\mathrm{ARIMA}(p - d, d, q) \times (P - D, D, Q)$ model for appropriate d and D with $0 \le 0 \le p$ and $0 \le D \le Q$. Assume that the observed values y_t of Y_t are available for the times $1 \le n \le N$, where $N > 0$ is a large enough integer. For the sake of simplicity, assume that $s \mid N$. We first discuss the seasonal part

$$(18.23) \qquad Y_t = \sum_{k=1}^{P} \Phi_k Y_{t-sk} + \epsilon_t - \sum_{k=1}^{Q} \Theta_k \epsilon_{t-s)}$$

of the model, where ϵ_t describes the error between times $t - s$ and t; what we mean by this error will be explained below. We determine the AR coefficients in this equation by using the least square approximation. That is, let Φ_k for k with $1 \le k \le P$ be such that

$$(18.24) \qquad \sum_{n=sP+1}^{N} \frac{1}{1 + y_n^2}\Big(y_n - \sum_{k=1}^{P} \Phi_k y_{n-sk}\Big)^2$$

is the least possible; see footnote 18.1 on p. 68 for an explanation of the reason to divide by $1 + y^2$. We determine the seasonal error vectors, we proceed similarly as we did around equation (18.5) except that now we need to take the seasons into account. Let $m \le N - sP - sQ$ be a large positive integer. Put

$$(18.25) \qquad \mathbf{y}_k = (y_{k-m+1}, y_{k-m+2}, \ldots, y_k)^T \qquad (k \ge m,\ s \mid k);$$

requiring $s \mid k$ is important here, so y_t and $y_{t'}$ occurs in the same component only if $t \equiv t' \mod s$. To determine the error vectors $\boldsymbol{\epsilon}_t$, orthogonalize the vectors $\mathbf{y}_t$ for t with $m \le t \le N$ and $s \mid t$. As before, the vectors $\mathbf{y}_k$ can be expressed as a linear combination of the error vectors as

$$(18.26) \qquad \mathbf{y}_t = \sum_{l=0}^{t-m} \psi_{l,t}\boldsymbol{\epsilon}_{t-sl} = \boldsymbol{\epsilon}_t + \sum_{l=1}^{t-m} \psi_{l,t}\boldsymbol{\epsilon}_{t-sl} \qquad (m \le t \le N,\ s \mid t),$$

similarly to equation (18.7). Proceeding similarly as we did after this equation, we arrive at an equation analogous to equation (18.13):

$$\Phi(B^s)Y_t = \Theta(B^s)\epsilon_t. \tag{18.27}$$

We expect that here all zeros of $\Theta(x)$ are outside the closed unit circle, while $\Phi(x)$ may have zeros on the unit circle, since we are creating an ARIMA model directly, rather than an ARMA model. Also note that the construction ensures that the constant term of $\Theta(x) = 1$, similarly as in equation (18.1). Expressing ϵ_T from equation (18.27, for a sequence of observations $\langle y_t : 1 \leq t \leq N \rangle$ we can then calculate the the approximate values $\hat{\epsilon}_t$ of the seasonal errors, similarly as we did in Subsection 9.5; for this, we need initial values for $\hat{\epsilon}_t$; the requirement that all zeros of $\Theta(x)$ are outside the unit circle ensure that the choice of the initial values of $\hat{\epsilon}_t$ do not significantly influence the values of $\hat{\epsilon}_t$ for moderately large t.

Next, we build an ARIMA model

$$\phi(B)\epsilon_t = \theta(B)e_t \tag{18.28}$$

using the sequence $\langle \hat{\epsilon}psilon_t : K \leq t \leq N \rangle$ as observed values; here $K > 0$ is used to discard the values of $\hat{\epsilon}t$ for small t for which the effect of the arbitrary choice of initial values cannot be considered small. Hence, we obtain

$$\phi(B)\Phi(B^s)Y_t = \phi(B)\Theta(B^s)\epsilon_t = \Theta(B^s)\phi(B)\epsilon_t = \Theta(B^s)\theta(B)e_t. = \theta(B)\Theta(B^s)e_t.$$

The first equation is obtained by multiplying equation (18.27) by $\phi(B)$ on the left; the second equation is uses the commutativity of polynomial multiplication, and the third equation follows from equation (18.28; the fourth equation again uses the commutativity of polynomial multiplication. That is, we have

$$\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)e_t.$$

This is the multiplicative seasonal ARIMA model we wanted to construct.

19 Bootstrap methods

In the paper in the paper [15] published in 1979, Bradley Efron described a number of statistical methods made feasible by the revolution in computing in the middle of the twenties century. Most statistical methods then in use, many of them still in use today, were invented in the early twentieth century were based on methods of computing that required relatively small amounts of calculation. Among these methods was *bootstrap*, Efron's own invention.

19.1 Bootstrap for independent identically distributed random variables

Bootstrap, as originally invented for independent, identically distributed random variables, can be described as follows. Assume we have a sample x_1, x_2, ..., x_n of measurements from a large population, so that these sample values can be regarded as values of a sequence of independent identically distributed random variables. Given the sample, we can estimate the population mean, but the question is how good this estimate is? Since we do not know anything about the distribution of the measurements, using normal distribution theory may lead to the wrong conclusion. In the bootstrap method we resample these measurements with replacement, we calculate the mean of each resample, and thereby we establish an empirical distribution of the means of the sample.

This can be described in a mathematical language as follows. Let $m > 0$ be an integer indicating the size of each resample, and let N be a large positive integer to indicate the number of resamples we want to create. For each i with $1 \le i \le N$ let

$$f_i : \{1, 2, \ldots, m\} \to \{1, 2, \ldots, n\}$$

be a random function (this function is not assumed to be one-to-one or onto). For each i, this will give a resample

$$x_{f_i(1)}, x_{f_i(2)}, \ldots, x_{f_i(m)}$$

of the original data. With

$$\mu_i = \frac{1}{m} \sum_{k=1}^{m} x_{f_i(k)}$$

we get a collection of the sample means μ_i. One can now devise a confidence interval for the population mean μ by choosing an a such that the about $5N/200 = .025N$ among the μ_i is less than a and choosing a b such that about $5N/200 = .025N$ among the μ_i is greater than b. Then one can say that $a \le \mu \le b$ with 95% confidence. The method can be used to set up estimates for other population parameters, such as the variance, median, etc.

19.2   Confidence intervals for multistep predictions in ARIMA models

Given an ARIMA model

(19.1) $$Y_t = \sum_{k=1}^{p} \phi_k Y_{t-k} + e_t - \sum_{k=1}^{q} \theta_k e_{t-k} \qquad (t \in \mathbb{Z}),$$

the errors e_t, also called *residuals*, can be estimated from an observed run of the time series, as described in Subsection 9.5[19.1] (as pointed out at the cited loaction, when using this method, a number of the beginning values of the residuals need to be discarded, because the choice of the initial values does not correspond to their actual values) or in Subsection 10.1. Assuming that Y_t has observed values y_t for t with $1 \le t \le N$, and the estimated values for the residuals is $\hat{e}_t$. Assume, further that the residuals are considered reliable for t with $K \le t \le N$. For $t > N$, an estimate for e_t is not available. For one prediction run, for $t > N$ one can define $\hat{e}_t$ and a randomly selected value from among the residuals $e_{t'}$ for t' with $K \le t' \le N$. In this way, replacing e_t with $\hat{e}_t$, one can use equation (19.1) repeatedly with $t = N + 1$, $N + 2$, ..., $N + k$ to predict Y_{N+k}. Making repeated predictions of Y_{N+k} with new random choices of the future residuals, one can construct an empirical distribution of Y_{N+k}, and using this empirical distribution, one can find a confidence interval for the predicted value of Y_{N+K}.

The application of this method relies on the tacit assumption that the residuals are independent identically distributed random variables. This assumption goes beyond the assumption of stationarity of the appropriately differenced time series used in the construction of the ARIMA model, since stationarity does only involves first and second moments, and says nothing about distributions. The assumption of strict stationarity would certainly imply this (see Subsection 5.1), but even without the assumption of strict stationarity one often makes this assumption about the residuals.

[19.1]That discussion concerned only ARMA models rather than ARIMA models. However, and ARIMA model is also an ARMA model for the appropriately differerced time series, with the same residuals.

19.3 Other applications of bootstrap for time series

Bootstrap methods have been extended from independent identically distributed random variables to other situations, and there are many other, more complicated applications of bootstrap methods for time series. See Kreiss and Lahiri [21], Politis [27], and Kirch and Politis [20]; the last one discusses bootstrap methods in the frequency domain.

20 The Fourier transform

20.1 The definition of the Fourier transform

Let f be a function on $\mathbb{R}$. Its Fourier transform is defined as

$$(20.1) \qquad \hat{f}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(y)e^{-ixy}\,dy,$$

assuming the integral exists. We then have

$$(20.2) \qquad f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{f}(y)e^{ixy}\,dy;$$

again f has to satisfy certain conditions for this integral to exist. The expression on the right-hand side is called the inverse Fourier transform. We will outline how to prove formula (20.2) while treating some convergence issues lightly.

Given a function f on $\mathbb{R}$ and a (large) integer N, we will represent f by a Fourier series on the interval $(-N\pi, N\pi)$. To do this, we write $y = x/N$

$$g(y) = f(Ny) = f(x),$$

and represent $g(y)$ by a Fourier series on $(-\pi, \pi)$ as [20.1]

$$(20.3) \qquad f(x) = g(y) = \sum_{n=-\infty}^{\infty} c_n e^{iny} = \sum_{n=-\infty}^{\infty} c_n e^{ixn/N} \qquad (-N\pi < x < N\pi),$$

where

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} g(y)e^{-iny}\,dy = \frac{1}{2N\pi} \int_{-N\pi}^{N\pi} f(x)e^{-ixn/N}\,dx \qquad (-\infty < n < \infty),$$

where the first equation holds according to equation (13.15), and the second equation was obtained by using the substitution $x = Ny$ and noting that then $g(y) = f(x)$. Writing

$$(20.4) \qquad h_N(t) = \frac{1}{2\pi} \int_{-N\pi}^{N\pi} f(x)e^{-ixt}\,dx,$$

we have $c_n = (1/N)h_N(n/N)$, and equation (20.3) becomes

$$(20.5) \qquad f(x) = \frac{1}{N} \sum_{n=-\infty}^{\infty} h_N(n/N)e^{ixn/N} \qquad (-N\pi < x < N\pi).$$

[20.1] If f is continuous and put together from finitely many monotonic pieces on finite intervals, the next equation will hold for all $y \in (-\pi, \pi)$, i.e., for all $x \in (-N\pi, N\pi)$, according to Dirichlet's theorem quoted in Subsection 13.2. Even then, it will not hold for $x = \pm N\pi$ unless $f(N\pi) = f(-N\pi)$. If f is not real-valued, Dirichlet's theorem can be applied separately to the real part and the imaginary part of f, assuming that those are continuous and put together from finitely many monotonic pieces on finite intervals.

where, as above, we wrote $x = y/N$. Putting

$$(20.6) \qquad h(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x)e^{-ixt}\, dx,$$

we have $\lim_{N \to \infty} h_N(t) = h(t)$ according to equation (20.4). Making $N \to \infty$ in (20.5), the sum approximates an integral, and h_N approaches h, and so we obtain

$$(20.7) \qquad f(x) = \int_{-\infty}^{\infty} h(t)e^{ixt}\, dt \qquad (-\infty < x < \infty).$$

Equations (20.6) and (20.7) are identical to equations (20.1) and (20.2) with $\hat{f}(x) = \sqrt{2\pi}\, h(t)$.

20.2 The Fourier transform is an isometry

The formula corresponding to Parseval's identity (13.16) is

$$(20.8) \qquad \|f\|^2 = \int_{-\infty}^{\infty} |f(x)|^2\, dx = \int_{-\infty}^{\infty} |\hat{f}(x)|^2\, dx = \|\hat{f}\|^2$$

assuming both integrals exist. This means that for $f \in L^2(-\infty, -\infty)$, the norm of f and $\hat{f}$ is the same; a transformation of normed vector spaces that preserves norms is called an isometry.[20.2] The statement described by this equation is called Plancherel's theorem. The Fourier transform as described in (20.1) does not exist for every $f \in L^2(-\infty, \infty)$, but every such function can be approximated by a sequence f_n of functions such that f_n converges to f in norm; in fact, we can take f_n to be a continuous function that is 0 outside a finite subinterval of $(-\infty, \infty)$.[20.3] Taking a sequence of functions f_n such that $f = \lim_{n \to \infty} f_n$ (convergence in norm), we can put

$$\hat{f} = \lim_{n \to \infty} \hat{f}_n.$$

The convergence here is assured, since $\|\hat{f}_n - \hat{f}_m\| = \|(f_n - f_m)^{\hat{}}\| = \|f_n - f_m\|$.

Plancherel's theorem ({20.8) can be extended to inner products:

$$\langle f, g \rangle = \int_{-\infty}^{\infty} \big(f(x)\big)^* g(x)\, dx = \int_{-\infty}^{\infty} (\hat{f}(x))^* \hat{g}(x)\, dx = \langle \hat{f}, \hat{g} \rangle.$$

This is immediate from the identity

$$(20.9) \qquad 4f^*g = |f + g|^2 - |f - g|^2 + i|if + g|^2 - i|if - g|^2.$$

See Problem 20.1 for the proof of this identity.

[20.2]More generally, an isometry is a transformation of metric spaces that preserves distances.

[20.3]Such a function can be called a continuous function with *compact* support. The support of a function is a set that contains all the points where a function is nonzero. We do not need the concept of compact sets at this point, it suffices to say that every bounded closed interval is a compact set. So, a function is said to have compact support if it is zero outside a bounded closed interval.

20.3 The Fourier transform and convolution

Given two functions f and g on $\mathbb{R}$, according to equation (20.1) and the second equation in (5.5), we have

$$
\begin{aligned}
(f * g)\hat{\ }(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(\tau)g(y-\tau)\,d\tau \right) e^{-ixy}\,dy \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(\tau)g(y-\tau)e^{-ix\tau}e^{-ix(y-\tau)}\,d\tau \right) dy \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(\tau)g(y-\tau)e^{-ix\tau}e^{-ix(y-\tau)}\,dy \right) d\tau \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(\tau)g(u)e^{-ix\tau}e^{-ixu}\,du \right) d\tau \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(u)e^{-ixu}\,du \int_{-\infty}^{\infty} f(\tau)e^{-ix\tau}\,d\tau \ = \ \sqrt{2\pi}\,\hat{g}(x)\hat{f}(x) = \sqrt{2\pi}\,\hat{f}(x)\hat{g}(x),
\end{aligned}
$$

(20.10)

where the fourth equation was obtained by making the substitution $u = y - \tau$. That is, the Fourier transform converts a convolution into a product.[20.4] Of course, there are conditions f and g must satisfy in order that the transformations performed in the equations above be permissible, but we omit any discussion of them. Besides, such a discussion can be done much more fruitfully with Lebesgue integration theory than with Riemann integration.

One can derive a similar relation between convolutions of sequences and the Fourier series formed by these sequences as coefficients. Indeed, if f and g are functions on $\mathbb{Z}$ and

$$
F(x) = \sum_{k=-\infty}^{\infty} f(k)e^{ikx} \qquad \text{and} \qquad G(x) = \sum_{k=-\infty}^{\infty} g(k)e^{ikx}
$$

then

(20.11)
$$
F(x)G(x) = \sum_{k=-\infty}^{\infty} (f * g)(k)e^{ikx}.
$$

This is certainly true if the series representing $F(x)$ and $G(x)$ are absolutely convergent, and it simply reflects the rule for multiplying two-way power series.[20.5] The above relation shows one aspect of the importance of the Fourier transform for analysis of time series. Linear filters or convolutions (see Subsection 5.6) are important for analyzing or processing time series, and their effects are much easier to study in the frequency domain, since products are much easier to understand than convolutions.

20.4 Frequency filtering

To screen out certain frequencies from an incoming signal has been a concern for radio engineers for a long time; for example, when you tune into a radio station, you do not want to listen to the neighboring station at the same time. This was accomplished by analog circuits, but today, a lot

---

[20.4]The factor $\sqrt{2\pi}$ on the right-hand side is somewhat of a nuisance. To avoid this, sometimes when discussion convolutions and Fourier transforms together, one puts a factor of $1/\sqrt{2\pi}$ in front of the integral in the definition of convolution in equation (5.5).

[20.5]If $z = e^{ix}$ then $e^{ikx} = z^k$, so the series representing F and G are two-way infinite power series.

of filtering is done by mathematically processing the time series obtained by sampling the analog signal. Practically, one would only be interested in the discrete series resulting by sampling rather than the continuous signal, that is, in the Fourier series rather than the Fourier transform, but, for a theoretical understanding of the issues studying the Fourier transform is very important. The isometry of the Fourier transform described in equation (20.8) is often easier the work with than the analogous Parseval identities (13.16) or (14.6). Mathematically, filtering out frequencies in the signal expressed by the function $f(t)$ amounts to taking the characteristic function $\chi_{[a,b]}$ of the interval,[20.6] and then taking the inverse Fourier transform of the function $\chi_{[a,b]}\hat{f}$. One can use functions other than characteristic functions of intervals for filtering. In fact, in analog processing the filtering function that can be realized will only approximately be the characteristic function of an interval. This kind of filtering is called filtering in the frequency domain. One can also use filtering in the time domain, or in spatial domain (for image transmission), or in the time-space domain.

20.5   Spectral analysis: what for?

The book [11, §9.6, pp. 183–190] describes a number of applications of frequency analysis. An especially interesting one concerns fault detection on electric motors. Electric motors vibrate, and the vibration has typical frequencies, and faults such as a broken rotor bars[20.7] changes these frequencies. Monitoring these frequencies can be used to detect faults. Spectral analysis can be used for stationary time series; vibration of electric motors naturally generate stationary time series – in the electric motor example, the signal was monitored 400 times a second, so the time series can safely be assumed to be stationary. Monitoring frequency variations in nonstationary time series can be accomplished with wavelets – see Section 22.

20.6   Problems

Problem 20.1. Prove equation (20.9).

Problem 20.2. Prove the analog of equation (20.9) for inner products and norms. That is, given a complex inner product $\langle \cdot, \cdot \rangle$ and the induced norm $\| \cdot \|$ in a vector space V over $\mathbb{C}$, show that

$$(20.12) \qquad 4\langle f, g \rangle = \|f + g\|^2 - \|f - g\|^2 + i\|if + g\|^2 - i\|if - g\|^2.$$

Problem 20.3. Let V be a normed vector space over $\mathbb{R}$. Show that the norm $\| \cdot \|$ is induced by an inner product if and only if

$$(20.13) \qquad \|f + g\|^2 + \|f - g\|^2 = 2\|f\|^2 + 2\|g\|^2 \qquad \text{for all} \quad f, g \in V.$$

This identity is called the *parallelogram identity*.[20.8] **Note:** This problem is difficult. The result is due to Maurice René Fréchet, John von Neumann, and Pascual Jordan.

Problem 20.4. Let V be a normed vector space over $\mathbb{C}$. Show that the norm $\| \cdot \|$ is induced by an inner product if and only if it satisfies equation (20.13).

Problem 20.5. Find the Fourier transform of

$$f(x) = e^{-\alpha^2 x^2},$$

where $\alpha > 0$.

[20.6]See footnote 17.7 on p. 67.

[20.7]A certain part of an alternating current induction motor – see [35]. Note that the example describes the alternating current frequency as 50 Hz, common in Europe. In the USA, the alternating current frequency is 60 Hz.

[20.8]If the vectors f and g stand for two sides of a parallelogram, the identity expresses the statement that the sum of squares of the diagonals of a parallelogram is equal to the sum of squares of the sides.

21   The Haar orthonormal system

In 1909, Alfred Haar designed an interesting system of orthonormal functions that became the focus of special interest in the light of later developments.[21.1] The functions χ_n are defined on the interval $[0,1]$ as follows. We put $\chi_0(x) = 1$ for all x with $0 \le x \le 1$. Further, we put

$$\chi_1(x) = \begin{cases} 1 & \text{if } 0 \le x < 1/2, \\ -1 & \text{if } 1/2 < x \le 1. \end{cases}$$

For each $n \ge 1$ divide the interval $[0,1]$ into 2^n equal intervals, and let these subintervals be denoted as I_n^j $(1 \le j \le 2^n)$. We write

$$\chi_n^{(k)}(x) = \begin{cases} 2^{(n-1)/2} & \text{inside } I_n^{2k-1}, \\ -2^{(n-1)/2} & \text{inside } I_n^{2k}, \\ 0 & \text{elsewhere in } (0,1) \end{cases} \qquad \text{for } 1 \le k \le 2^{n-1}.$$

At the jumps in the interval $(0,1)$, the value of $\chi_n^{(k)}$ will be the arithmetic mean of its values in the neighboring intervals. Finally, we define $\chi_n^{(k)}(x)$ for 0 and 1 for it to be continuous at these points ($n \ge 2$ and $1 \le k \le 2^{n-1}$) in the interval $[0,1]$.

As we will see below, the Haar system is a complete orthonormal with respect to the inner product

$$\langle f, g \rangle = \int_0^1 \big(f(x)\big)^* g(x), dx.$$

Nothing is really gained by considering complex-valued function, so, for the sake of simplicity, we may assume that f and g are real valued, and then the complex conjugation in this formula may be omitted.

We used Haar's original notation for these functions. To simplify the notation, write $\chi_0^{(0)} = \chi_0$ and $\chi_1^{(1)} = \chi_1$. The normality of the Haar system, i.e., that

$$\int_0^1 |\chi_n^{(k)}(x)|^2 \, dx = 1,$$

is easy to see. Further, we can also see that

$$\int_0^1 \chi_n^{(k)}(x) \, dx = 0 \qquad \text{for} \quad n \ge 1.$$

From this, the orthogonality

$$\int_0^1 \chi_n^{(k)}(x) \chi_{n'}^{(k')}(x) \, dx = 0 \qquad \text{if} \quad (n,k) \ne (n',k')$$

also follows. Indeed, if $n = n'$, then at least one of $\chi_n^{(k)}(x)$ and $\chi_n^{(k')}(x)$ is zero, with the exception of at most a single value of x. If $n < n'$, then $\chi_n^{(k)}$ is constant on the union the the two intervals $I_{n'}^{2k'-1}$ and $I_{n'}^{2k'}$ where $\chi_{n'}^{(k')} \ne 0$, except perhaps at the endpoints of the union.

We formulate the completeness as a separate theorem:

[21.1]Haar's original paper appeared in 1910, but on the first page it says that it is essentially an unchanged version of his "Göttinger Inauguraldissertation," that is, the dissertation written to obtain habilitation at the University of Göttingen, Germany. See footnote 13.6 on page 46 concerning habilitation.

Theorem 21.1. *The Haar system is a complete orthonormal system in $L^2[0,1]$.*

We have already established orthonormality. For the proof of completeness, we will use Lemma 17.2 on p. 66. For this, note that the (finite) linear combinations of the characteristic functions of the intervals I_n^k ($n \geq 1$, $1 \leq k \leq 2^n$) form a dense set in $L^2[0,1]$.

One really needs to study Lebesgue integration to really understand this statement, if for no other reason that the definition of $L^2[0,1]$ relies on Lebesgue integration. For an intuitive understanding for those only familiar with Riemann integration, note that every Riemann integrable function can be approximated by a step function with partition points coming from among the endpoints of the intervals I_n^k; approximation is meant here in the sense that the integral of the absolute value of the difference is small. To appreciate the difference between the Riemann integral and the Lebesgue integral, consider the function f on $[0,1]$ that is 1 at rational points and 0 elsewhere. This function is not Riemann integrable; its Lebesgue integral is 0.

In probability theory, this property of the Riemann integral should be understood to be a deficiency. Consider, for example, a random variable X with values uniformly distributed in the interval $[0,1]$, and ask the question: what is the probability of X assuming a rational value. As there are only countably many rational numbers, and the probability of $X = r$ for any specific r is zero, the probability of X being rational is 0, because of the σ-additivity axiom of probability theory, saying that if A_n are mutually exclusive events ($1 \leq n < \infty$) and A is the event that at least one of the A_n will occur, then

$$P(A) = \sum_{n=1}^{\infty} P(A).$$

This property of σ-additivity is built into the definition of Lebesgue integral, whereas it is not in that of the Riemann integral.

Proof of Theorem 21.1. Given $m \geq 2$, let D_m be the subspace of $L^2[0,1]$ spanned by the (finite) linear combinations of the characteristic functions of I_n^k for $1 \leq n \leq m$ and and $1 \leq k \leq 2^n$.

The dimension of this space of 2^m; the reason for this is that for $n < m$ and for $1 \leq k' \leq 2^{n'}$ the characteristic function of $I_n^{k'}$ is a linear combination of the characteristic functions I_m^k, $1 \leq k \leq 2^m$.[21.2] This is exactly the number of Haar functions $\chi_n^{(k)}$ belonging to this space, i.e., $\chi_n^{(k)}$ for $n = k = 0$ and for $1 \leq n \leq m$ and $1 \leq k \leq 2^m$, since

$$1 + \sum_{n=1}^{m} 2^{n-1} = 2^m.$$

Therefore these Haar functions span D_m. Therefore the Haar functions span $D = \bigcup_{m=1}^{\infty} D_m$. As D is dense in $L^2[0,1]$, the Haar system is complete by Lemma 17.2. □

21.1 Frequency filtering of the Haar system

In time series with changing characteristics, one wants to filter out distant parts; that is, one wants to filter in the time domain. In this sense, the Haar system (when adapted to discrete time series) is excellent, since it perfectly filters out far-away effects. Its frequency performance, however, is

---

[21.2]Observe that the value of these functions at the end points of the intervals $i_n^{(l)}$ make no difference. The real reason for this is that $L^2[0,1]$ is a space of equivalence classes of functions, and not a space of functions, in spite of one saying the opposite in loose parlance. See the second paragraph of footnote 17.4 on 65.

another story. Indeed, for the Fourier transform of $\chi_n^{(k)}$ for $x \neq 0$ we have

$$
\begin{aligned}
\hat{\chi}_n^{(k)}(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \chi_n^{(k)}(x) e^{-ixy}\, dy \\
&= \frac{1}{\sqrt{2\pi}} \left(\int_{(2k-2)/2^n}^{(2k-1)/2^n} - \int_{(2k-1)/2^n}^{2k/2^n} \right) 2^{(n-1)/2}\, e^{-ixy}\, dy \\
&= \frac{2^{(n-1)/2}}{-ix\sqrt{2\pi}} \left(-e^{-i(2k-2)/2^n} + 2e^{-i(2k-1)/2^n} + e^{-2ik/2^n} \right) = O\left(\frac{1}{x} \right)
\end{aligned}
$$

as $x \to \infty$; the last equation ignores the dependence on n.[21.3] The problem here is that $1/x$ does not tend to zero fast enough when $x \to \infty$. In engineering terms, as says that the functions of the Haar system have poor performance in frequency filtering.

22 Wavelets

22.1 Haar wavelet and multiresolution analysis

Let $V = L^2(\mathbb{R})$ (the real L^2 space), and $I_{n,k}$ be the interval $\left[(k-1)2^n, k2^n \right)$ for $\infty < k < \infty$. Let

(22.1) $$ V_n = \{ f \in V : f \text{ is constant on } I_{n,k} \text{ for each } k \in \mathbb{Z} \}. $$

Let $\phi = \chi_{[0,1)}$ be the characteristic function of the interval $[0,1) = I_{0,0}$, that is,

$$
\phi(x) = \begin{cases} 1 & \text{if } 0 \le x < 1, \\ 0 & \text{otherwise.} \end{cases}
$$

and let

$$
\psi(x) = \begin{cases} 1 & \text{if } 0 \le x < 1/2, \\ -1 & \text{if } 1/2 \le x < 1, \\ 0 & \text{otherwise.} \end{cases}
$$

The function ϕ is called the Haar scaling function, and ψ, the Haar mother wavelet.[22.1] We put

$$
\phi_{n,k}(x) = 2^{-n/2} \phi(2^{-n}x - k),
$$
$$
\psi_{n,k}(x) = 2^{-n/2} \psi(2^{-n}x - k)
$$

for $n, k \in \mathbb{Z}$. That is,

$$
\phi_{n,k}(x) = \begin{cases} 2^{-n/2} & \text{if } x \in I_{n,k}, \\ 0 & \text{otherwise.} \end{cases}
$$

[21.3]The "big Oh" and "little oh" symbols were introduced by Edmund Landau. The symbols are very convenient, but often their exact meaning must be ascertained from the context. Given a function $f(x)$, which is usually, but not necessarily assumed to be positive, the symbol $O\big(f(x)\big)$ denotes a function $g(x)$ such that $g(x)/f(x)$ remains bounded when $x \to a$, or $x \searrow a$, or $x \nearrow a$ (i.e., x tends to a from the right, or from the left), where usually $a = +\infty$, or $a = -\infty$, or $a = \pm\infty$ or $a = 0$, or else a is any other value; the value of a and how it is approached should be understood from the context. Similarly, $o\big(f(x)\big)$ denotes a function $g(x)$ such that $\lim_{x \to a} f(x)/g(x)$, (or $\lim_{x \searrow a} f(x)/g(x)$, or $\lim_{x \nearrow a} f(x)/g(x)$), where, again a and how it is approached should be understood from the context.
[22.1]The scaling function is occasionally called the father wavelet.

and

$$\psi_{n,k}(x) = \begin{cases} 2^{-n/2} & \text{if } x \in I_{n-1,2k-1}, \\ -2^{-n/2} & \text{if } x \in I_{n-1,2k}, \\ 0 & \text{otherwise.} \end{cases}$$

For fixed n, The system of functions $\phi_{n,k}$ is orthonormal in V_n. The orthogonality is clear, since if $k \neq k'$ then for all $x \in \mathbb{R}$, one of $\phi_{n,k}(x)$ and $\phi_{n,k'}(x)$ is zero. The Haar system can be expressed in terms of these functions as[22.2]

$$\begin{aligned} \chi_0 &= \phi, \\ \chi_1 &= \psi, \\ \chi_n^{(k)} &= \psi_{-n+1,k} \qquad \text{for } n \geq 2 \text{ and } 1 \leq k \leq 2^{n-1}. \end{aligned}$$

It is also clear that any function f_n in V_n can be expressed as a sum

(22.2)
$$f_n(x) = \sum_{k=-\infty}^{\infty} c_{n,k}\, \phi_{n,k}(x),$$

where c_k is the constant value of $2^{n/2} f(x)$ for $x \in I_{n,k}$.[22.3] This equation implies that the the system of functions $\phi_{n,k}$ is also complete in V_n (cf. Lemma 17.2 with $M = V$ in that Lemma – i.e., $M = V_n$ in the present case).

Observe that

(22.3)
$$\phi_{n+1,k} = \frac{1}{\sqrt{2}} \left(\phi_{n,2k-1} + \phi_{n,2k} \right)$$

and

(22.4)
$$\psi_{n+1,k} = \frac{1}{\sqrt{2}} \left(\phi_{n,2k-1} - \phi_{n,2k} \right).$$

Hence, the above equation becomes

$$\begin{aligned} f_n &= \sum_{k=-\infty}^{\infty} c_{n,k}\, \phi_{n,k} = \sum_{k=-\infty}^{\infty} \left(c_{n,2k-1}\, \phi_{n,2k-1} + c_{n,2k}\, \phi_{n,2k} \right) \\ &= \sum_{k=-\infty}^{\infty} \left(\frac{c_{n,2k-1} + c_{n,2k}}{2} \left(\phi_{n,2k-1} + \phi_{n,2k} \right) + \frac{c_{n,2k-1} - c_{n,2k}}{2} \left(\phi_{n,2k-1} - \phi_{n,2k} \right) \right) \\ &= \sum_{k=-\infty}^{\infty} \left(\frac{c_{n,2k-1} + c_{n,2k}}{\sqrt{2}} \phi_{n+1,k} + \frac{c_{n,2k-1} - c_{n,2k}}{\sqrt{2}} \psi_{n+1,k} \right) \end{aligned}$$

[22.2]Note quite, since we equality at the end points of the intervals $I_{n,k}$ is not guaranteed. However, these functions are still equal a.e., i.e., in the sense of L^2. See footnote 21.2 on page 81.

[22.3]As a consequence of orthonormality, we must have

$$c_{n,k} = \int_{I_{n,k}} \phi_{n,k}(x) f(x)\, dx = \langle \phi_{n,k}, f \rangle.$$

We did not use complex conjugate in this equation, since f is assumed to be real valued. It is also easy to check this equation directly.

Writing

$$(22.5) \qquad f_{n+1} = \sum_{k=-\infty}^{\infty} \frac{c_{n,2k-1} + c_{n,2k}}{\sqrt{2}} \phi_{n+1,k}$$

and

$$(22.6) \qquad g_{n+1} = \sum_{k=-\infty}^{\infty} \frac{c_{n,2k-1} - c_{n,2k}}{\sqrt{2}} \psi_{n+1,k},$$

we have $f_n = f_{n+1} + g_{n+1}$. Further, $f_{n+1} \in V_{n+1}$. and $\langle f_{n+1}, g_{n+1} \rangle = 0$. This latter equation holds since $\langle \phi_{n+1,k}, \psi_{n+1,k'} \rangle = 0$.[22.4] Write

$$W_{n+1} = \left\{ \sum_{k=-\infty}^{\infty} a_k \psi_{n+1,k} : \sum_{k=-\infty}^{\infty} |a_k|^2 < \infty \right\}.$$

It is worth restating this discussion in a more formal framework. We need a definition for this.

Definition 22.1. Let X and Y be subspaces of the inner product space U such that for each $x \in X$ and $y \in Y$ we have $\langle x, y \rangle = 0$. We then call X and Y orthogonal and we write

$$X \oplus Y = \{x + y : x \in X \quad \text{and} \quad y \in Y\}.$$

$X \oplus Y$ is called the orthogonal sum of the spaces X and Y. If $U = X \oplus Y$, then we can also write $Y = U \ominus X$.[22.5] If $U = X \oplus Y$ and $u \in U$, then the unique y for which $u = x + y$ is called the projection of u onto Y. The function P for which $Pu = y$ is called the projection operator onto Y.[22.6]

We have

$$(22.7) \qquad V_n = V_{n+1} \oplus W_{n+1} \qquad \text{for all} \quad n \in \mathbb{Z}.$$

Let P_n be the projection operator from V_n to V_{n+1}; we have $P_n f_n = f_{n+1}$. There are further notable properties of the spaces involved that will be important for describing a more general setting of multiresolution analysis. We have

$$(22.8) \qquad V_{n+1} \subset V_n \qquad \text{for all} \quad n \in \mathbb{Z}.$$

Further,

$$(22.9) \qquad \bigcap_{n=-\infty}^{\infty} V_n = \{0\};$$

[22.4]If $k \neq k'$ then for any $x \in \mathbb{R}$, one of $\phi_{n+1,k}(x)$ and $\psi_{n+1,k'}$ is zero.

[22.5]One needs to be a little careful here. The symbol $\oplus$ is also used to indicate the direct sum of two vector spaces. When X and Y are subspaces of a vector space U (no inner product is assumed here), and $X \cap Y = \{0\}$, then the direct sum of X and Y is defined as

$$X \oplus Y = \{x + y : x \in X \quad \text{and} \quad y \in Y\}.$$

There is no real conflict here, but there is one important difference. If U is the orthogonal sum of X and Y, then knowing U and X, we can find Y. On the other hand, if U is only the direct sum of X and Y, then knowing X, we can have several choices for Y.

[22.6]It is customary to write Pu instead of $P(u)$. This is a general custom for operators. P is a linear operator (also called a linear transformation in linear algebra).

This is because a function belonging to all the spaces V_n has to be constant on the intervals $[0, 2^n)$ and $[-2^n, 0)$ for all n, and then it also needs to be square integrable. We also have

$$(22.10) \qquad \mathrm{cl}\left(\bigcup_{n=-\infty}^{\infty} V_n \right) = L^2(\mathbb{R});$$

here $\mathrm{cl}(U)$ denotes the *closure* of U. For this closure to make sense, U needs to be a subspace of a given normed space, say Z[22.7] (at present, this space is V). A subspace Z is closed if given any sequence of elements of Z that is convergent in norm then the limit of this sequence is also in Z. The closure of U is the smallest closed subspace of Z that includes U.[22.8] Finally, we have

$$(22.11) \qquad \bigoplus_{n=-\infty}^{\infty} W_n = L^2(\mathbb{R});$$

the symbol on the left indicates the closure of the subspace formed by all finite linear combinations of vectors in $\bigcup_{n=-\infty}^{\infty} W_n$, and the direct sum sign also indicates that the subspaces W_n are pairwise orthogonal. The equation is the consequence of equations of equations (22.7), (22.8), (22.9), and (22.10), as we will see next.

Write

$$f_{n+1}(x) = \sum_{k=-\infty}^{\infty} c_{n+1,k}\, \phi_{n+1,k}(x),$$

and

$$g_{n+1}(x) = \sum_{k=-\infty}^{\infty} d_{n+1,k}\, \psi_{n+1,k}(x).$$

Equations (22.2), (22.5), and (22.6) imply

$$(22.12) \qquad c_{n+1,k} = \frac{c_{n,2k-1} + c_{n,2k}}{\sqrt{2}} \quad \text{and} \quad d_{n+1,k} = \frac{c_{n,2k-1} - c_{n,2k}}{\sqrt{2}}.$$

Since we have $f_n = f_{n+1} + g_{n+1}$ for all n, equation (22.11) implies that for any $f_n \in V_n$ we have

$$(22.13) \qquad f_n = \sum_{m=n+1}^{\infty} \sum_{k=-\infty}^{\infty} d_{m,k} \psi_{m,k}.$$

This equation together with equation (22.10) implies (22.11). Equations (22.12) will point to way to compute the coefficients in the wavelet expansion of a function. This will be further elaborated below in a discussion of the discrete wavelet transform.

The above discussion is based mainly on [14, Chapter 5, pp. 129–].[22.9] This is an award-winning book, but it has some prerequisites in functional analysis and harmonic analysis to read it.

[22.7]An inner product space is also a normed space with the norm induced by the inner product, as we pointed out above.

[22.8]That is, $\mathrm{cl}(U)$ consists of the limits of all sequences convergent in norm whose elements come from U.

[22.9]The mistakes are mine.

22.2 What are wavelets?

Given a function $\psi \in L^2(\mathbb{R})$, called the mother wavelet, and a function $f \in L^2(\mathbb{R})$, the continuous wavelet transform is[22.10]

$$F_w(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(x) \left(\psi \left(\frac{x-b}{a} \right) \right)^* dx \qquad (a, b \in \mathbb{R} \text{ and } a > 0).$$

Often, one only considers discretely labeled wavelets, meaning that $F_w(a,b)$ is only considered for certain discrete values of a and b.[22.11] An illustration for this were the Haar wavelets in Subsection 22.1. Orthonormality, as exemplified by the Haar wavelets is a useful property in allowing efficient computer algorithms. There are also non-orthonormal wavelets that retain most of the computational advantages of orthonormal wavelets.

Wavelets are used for localized frequency analysis of data. For time series occurring in practice this is very important, since time series are usually non-stationary, and Fourier analysis is applicable only to stationary time series. The short-time Fourier transform considers only a part of the time series to keep track of frequency changes. It is used for analysis of a fixed frequencies at a fixed bandwidth (the difference between the upper and lower frequencies in the analyzed range of frequencies). Wavelets automatically adapt the analyzed frequency range to the size of the frequency.

Restricting time series both in time and frequency is mathematically impossible, since one needs infinitely long time to measure a frequency exactly. This is related to the Heisenberg's uncertainty relations in physics concerning the determination of the location and the momentum (velocity times mass) of a particle.[22.12] Daubechies [14, §2.3, pp. 21–23] discusses the example of a phone conversation, which is of finite time, and also of limited bandwidth, since the phone line is capable of transmitting frequencies only in a certain range. So, how well can a function be represented under such circumstances. The problem is not an easy mathematical problem, and its solution involves eigenvalues and eigenfunctions of integral and differential operators.

22.3 Smoothness and frequency filtering

As we discussed in Subsection 21.1, the frequency filtering performance of the Haar wavelet is poor; the main reason for this is the sharp discontinuity of the Haar wavelet. In order to get better performance, one needs smooth wavelets. The reason smoothness help frequency filtering can be seen by integration by parts. Indeed, assume $\psi(x) = 0$ outside a bounded interval (such a function is called compactly supported; see footnote 20.3 on p. 77.) Assume, further, that ψ is continuously differentiable; this will allow integration by parts. We have

(22.14)
$$\hat{\psi}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi(y) e^{-ixy} \, dy = \frac{1}{-ix\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi'(y) e^{-ixy} \, dy$$

for $x \neq 0$;[22.13] note that there is no integrated-out term, since $\psi(y) = 0$ for large y. The x in the denominator of the factor on the right-hand side indicates the speed of convergence of $\hat{\psi}(x) \to 0$ as $x \to 0$.[22.14] If ψ is continuously differentiable more than once, then we can repeat integration

---

[22.10]There is a technical condition, called the admissibility condition, that need to be imposed on ψ in order that f can be reconstructed from its wavelet transform. See [14, Section 1.3, p. 7].

[22.11]Discretely labeled wavelets are to be distinguished from the discrete wavelet transform, discussed below.

[22.12]This relation is certainly not perfect. For time series, the statement is a mathematical result, for physics, it is a basic principle that supports arguments even in cases when the exact equations governing a physical system are not known.

[22.13]We have $i = 1/(-i)$, so we could simplify the right-hand side a little, but that is beside the point.

[22.14]Without the x in the denominator, one would expect a rate of convergence of $O(1/x)$, as in the Haar wavelet – see Subsection 21.1. This factor indicates that the rate of convergence is at least $O(1/x^2)$.

by parts to show even better frequency filtering. Given any positive integer n, in [13], Daubechies developed a method to construct orthonormal wavelets that are zero outside a bounded interval and are continuously differentiable n times.

22.4   A short history of wavelets

22.5   Wavelets and image analysis

Given a complete orthonormal system of wavelets $\psi_{n,k}(x)$, one can construct a two-dimensional complete orthonormal system of wavelets if two dimensions by taking the system of functions

$$\Psi_{n,k,n',k'}(x,y) = \psi_{n,k}(x)\psi_{n',k'}(y).$$

Such a system of two-dimensional wavelets can be called the *tensor product* of one-dimensional wavelets. A more interesting scheme of producing two-dimensional wavelets is using multiresolution analysis in two-dimensions directly; that is, the method the basic features of which were described in Subsection 22.1, can be generalized to two dimensions without relying on one-dimensional wavelets; see [14, Chapter 10, pp. 313–]. Smoothness of two-dimensional wavelets is important for avoiding edge effects, caused by sharp jumps in the Haar wavelet. The site [1] has nice pictures showing the wavelet decomposition of images, and illustrates various uses (such as e.g. edge detection) of wavelets with pictures.

22.6   The discrete Haar wavelet transform

Let

$$f_0(x) = \sum_{k=-\infty}^{\infty} c_{0,k}\,\phi_{0,k}(x),$$

Start with a finite sequence of the coefficients $c_{0,k}$; these are perhaps the sampled value of a continuous time series (the coefficients outside the sampled range can taken to be zero). Using equations (22.12), and can calculate the coefficients $c_{n,k}$ and $d_{n,k}$ for $n > 0$. These equations show the number of coefficients $c_{n+1,k}$ in the nonzero range is half of the number of coefficients $c_{n,k}$ in the nonzero range. Similarly, the number of coefficients $d_{n+1,k}$ in the nonzero range is half of the number of coefficients $c_{n,k}$ in the nonzero range. So, after a while, all coefficients will be zero. The coefficients $d_{m,k}$ for $m > 0$ and $z \in \mathbb{Z}$ will be the coefficients of the wavelet expansion of f_0; cf. equation (22.13).

22.7   Orthogonal wavelets

We will generalize the framework described in Subsection 22.1 on account of the Haar wavelets. The starting point again will be the selection of two functions ϕ, called the scaling function, and ψ, called the mother wavelet. These two functions will determine the wavelets to be constructed, and how to select these functions is a difficult problem. These two functions, and the subspaces V_n and W_n constructed with the aid of them will satisfy properties analogous to those described in Subsection 22.1.

   The closed subspaces[22.15] V_n and W_n of $L^2(\mathbb{R})$ will satisfy equations (22.7)–(22.11). Equation (22.1) will be no longer in force – that equation applied only to the Haar wavelets. Instead, we will

---

[22.15]A subspace U is closed if $\mathrm{cl}(U) = U$. See footnote 22.8 on p. 85.

require that

$$(22.15) \qquad \begin{aligned} V_n = \{g : &\text{there is an } f \in V_0 \text{ and a } k \in \mathbb{Z} \text{ such that} \\ &\text{for all } x \in \mathbb{R} \text{ we have } g(x) = f(2^{-n}x - k)\} \end{aligned}$$

for all $n \in \mathbb{Z}$. Note that for $n = 0$ this means that if $f \in V_0$ and $g(x) = f(x - k)$ for some integer k and for all reals x then we also have $g \in V_0$. Assume $\phi \in V_0$ and $\psi \in W_0$, and for each $n, k \in \mathbb{Z}$, put

$$(22.16) \qquad \begin{aligned} \phi_{n,k}(x) &\overset{def}{=} 2^{-n/2}\phi(2^{-n}x - k), \\ \psi_{n,k}(x) &\overset{def}{=} 2^{-n/2}\psi(2^{-n}x - k). \end{aligned}$$

Assume that $\{\phi_{n,k} : k \in \mathbb{Z}\}$ is an orthonormal basis of V_n and $\{\psi_{n,k} : k \in \mathbb{Z}\}$ is an orthonormal basis of W_n.[22.16] We have[22.17]

$$(22.17) \qquad \phi(x) = \sum_{k=-\infty}^{\infty} h_k \sqrt{2}\, \phi(2x - k) \qquad \text{for a.e. } x$$

with some numbers h_k, because $\phi(x) \in V_0 \subset V_{-1}$, and the functions $\sqrt{2}\,\phi(2x - k) = \phi_{-1,k}(x)$ form an orthonormal basis of V_{-1}. Similarly,

$$(22.18) \qquad \psi(x) = \sum_{k=-\infty}^{\infty} g_k \sqrt{2}\, \phi(2x - k) \qquad \text{for a.e. } x$$

with some numbers g_k, because $\psi(x) \in W_0 \subset V_{-1}$.

Assuming ϕ and ψ are continuous, these equations hold everywhere. In this case, the values of ϕ at the integers determine the values of ϕ and ψ at places $m/2$ for all integers m. The values of ϕ at these points then determine values of ϕ and ψ at all points $m/4$ for integer m. Repeating this argument, we can see that the values of ϕ at integers determine the values of ϕ and ψ at all dyadic rationals.[22.18] Then, by continuity, $\phi(x)$ and $\psi(x)$ are determined for all x.

In many cases, all but finitely many of the coefficients are zero in equations (22.17) and (22.18). This is certainly true if both ϕ and ψ are zero outside a finite interval (i.e., when ϕ and ψ have compact support.[22.19] The method to construct compactly supported smooth wavelets was invented by Daubechies, and it was described in [13] and also in [14]. It involves very sophisticated mathematics using the Fourier transform, estimating products of certain trigonometric polynomials,[22.20] eigenvalues, and polynomial algebra. Compactly supported smooth wavelets are indispensable for storing pictures on your cellphone. It is interesting to reflect on the abstract mathematical tools needed to develop such ubiquitous applications. In a book first published in 1940, G. H. Hardy [18]

[22.16]Here basis is meant in the the sense of normed vector space. That is, every element of the vector space can be represented as an *infinite* linear combination of the basis vectors. Linear independence is still meant in the sense of finite linear combinations – though this is not an issue, since linear independence is a consequence of orthonormality.

We need to make these assumptions only in case $n = 0$, when they in effect define the spaces V_0 and W_0 in terms of the functions ϕ and ψ, respectively. For other values of n, they care consequences of equations (22.16), (22.15), and (22.7), as one can see after some consideration.

[22.17]As indicated, these equations hold for almost every (a.e.) x. See the comment in the second paragraph of footnote 17.4 on p. 65.

[22.18]A *dyadic rational* is a number $m/2^n$ for all integers $n > 0$ and m.

[22.19]If ψ has compact support, then the wavelet is called compactly supported. If ϕ has compact support, then it follows that ψ also has compact support; the proof of this is, however, technical. See [14, Section 6.1, p. 167].

[22.20]Called Riesz products, named after F. Riesz, who was mentioned in footnote 17.4 on 65.

reflected on the practical usefulness of mathematics, and tried to draw the boundary between pure and applied mathematics. The boundary has considerably shifted since then for many reasons; the invention of computers played a major role, making vast segments of pure mathematics useful in applications. The book is an amusing light read.

22.8 The discrete wavelet transform

Given any integer m, using equations (22.16), we can rewrite equations (22.17) and (22.18) can be rewritten as

$$\phi_{0,m}(x) = \phi(x-m) = \sum_{k=-\infty}^{\infty} h_k \sqrt{2}\, \phi\big(2(x-m)-k\big) = \sum_{k=-\infty}^{\infty} h_k \phi_{-1,2m+k}(x) \qquad \text{for a.e. } x,$$

$$\psi_{0,m}(x) = \psi(x-m) = \sum_{k=-\infty}^{\infty} g_k \sqrt{2}\, \phi\big(2(x-m)-k\big) = \sum_{k=-\infty}^{\infty} g_k \phi_{-1,2m+k}(x) \qquad \text{for a.e. } x.$$

Using equations (22.16) again, these equations imply that for any integers m and n we have

(22.19)
$$\phi_{n+1,m} = \sum_{k=-\infty}^{\infty} h_k \phi_{n,2m+k},$$
$$\psi_{n+1,m} = \sum_{k=-\infty}^{\infty} g_k \phi_{n,2m+k},$$

where the equation of functions is meant a.e.

We want to express the functions on the right-hand side in terms of the functions on the left-hand side. Orthonormality makes this easy. Indeed, we have

$$\langle \phi_{n+1,m}, \phi_{n,l} \rangle = \left\langle \sum_{k=-\infty}^{\infty} h_k \phi_{n,2m+k}, \phi_{n,l} \right\rangle = \sum_{k=-\infty}^{\infty} h_k^* \langle \phi_{n,2m+k}, \phi_{n,l} \rangle$$

$$= \sum_{k=-\infty}^{\infty} h_k^* \,\delta_{2m+k,l} = \sum_{k=-\infty}^{\infty} h_k^* \,\delta_{k,l-2m} = h_{l-2m}^*.$$

The asterisk here indicates complex conjugation. Similarly,

$$\langle \psi_{n+1,m}, \phi_{n,l} \rangle = g_{l-2m}^*.$$

Usually, both functions ϕ and ψ are real, in which case the coefficients h_k and g_k are real, and the complex conjugation can be omitted.[22.21] Since the functions $\phi_{n+1,m}$ and $\phi_{n+1,m}$ form an orthonormal basis of the space $V_{n+1} \oplus W_{n+1} = V_n$, according to equations (17.4) and (17.5) this means that

(22.20)
$$\phi_{n,l} = \sum_{m=-\infty}^{\infty} h_{l-2m}^* \, \phi_{n+1,m} + \sum_{m=-\infty}^{\infty} g_{l-2m}^* \, \psi_{n+1,m}.$$

[22.21]If $A = (a_{kl})$ is an $m \times n$ matrix with complex entries, then its *Hermitian conjugate* (also called *conjugate transpose*) is the matrix A^* is the $n \times m$ matrix with a_{kl}^* being the entry in the lth *row* and kth *column*. That is, after taking the transpose of A, we take the complex conjugate of each entry. A square matrix U is a unitary matrix if $U^*U = I$, where I is the identity matrix of the appropriate size. That is, $U^{-1} = U^*$. The coefficient matrix in the system (22.19) of equations can easily seen to be an orthonormal matrix.

A matrix is unitary if and only if its columns (or its rows) form an orthonormal system of vectors. The real unitary matrix is an orthogonal matrix. Orthogonal matrices were defined in Subsection 8.2.

Given $f \in V_n$, we can write f as

$$f = \sum_{l=-\infty}^{\infty} c_{n,l}\, \phi_{n,l} = \sum_{m=-\infty}^{\infty} \Big(\sum_{l=-\infty}^{\infty} c_{n,l}\, h_{l-2m}^* \Big) \phi_{n+1,m} + \sum_{m=-\infty}^{\infty} \Big(\sum_{l=-\infty}^{\infty} c_{n,l}\, g_{l-2m}^* \Big) \psi_{n+1,m},$$

where the second equation was obtained by using (22.20). As $f \in V_n = V_{n+1} \oplus W_{n+1}$, we can also write

$$f = \sum_{m=-\infty}^{\infty} c_{n+1,m}\, \phi_{n+1,m} + \sum_{m=-\infty}^{\infty} d_{n+1,m}\, \psi_{n+1,m}.$$

Since this representation is unique, comparing the last two displayed equation, we obtain

(22.21)
$$c_{n+1,m} = \sum_{l=-\infty}^{\infty} c_{n,l}\, h_{l-2m}^*,$$

(22.22)
$$d_{n+1,m} = \sum_{l=-\infty}^{\infty} c_{n,l}\, g_{l-2m}^*.$$

Let

$$f_0(x) = \sum_{k=-\infty}^{\infty} c_{0,k}\, \phi_{0,k}(x).$$

As in Subsection 22.6, start with a finite sequence of the coefficients $c_{0,k}$; these are perhaps the sampled value of a continuous time series (the coefficients outside the sampled range can taken to be zero). Using (22.21) and (22.22) we can calculate the coefficients for $n > 0$. Assuming only a finite number of the coefficients h_k and g_k are nonzero, at for each n the number of coefficients gets approximately halved. Hence we will find an $N \in \mathbb{Z}$ such that all coefficients $c_{n,k}$ and $d_{n,k}$ will be zero for $n > N$. The coefficients $d_{m,k}$ for $m > 0$ and $z \in \mathbb{Z}$ will be the coefficients of the wavelet expansion of f_0:

$$f_0 = \sum_{m=1}^{\infty} \sum_{k=-\infty}^{\infty} d_{m,k} \psi_{m,k}.$$

In the language of electric engineering, equation (22.21) represents a low-pass filter, i.e., filtering out (discarding) high frequencies, that is, the finer features of the signal (those represented by elements of the space W_{n+1}), and equation (22.22) represents a high-pass filter, i.e., filtering out (discarding) low frequencies, that is, the cruder features of the signal (those represented by elements of the space V_{n+1}).

The Haar wavelets fit into this pattern as follows. Comparing equations (22.5), (22.6), and (22.19), we can see that

$$h_{-1} = h_0 = \frac{1}{\sqrt{2}}, \quad g_{-1} = \frac{1}{\sqrt{2}}, \quad g_0 = -\frac{1}{\sqrt{2}},$$

and $h_k = g_k = 0$ for $k \in \mathbb{Z}$ with $k \neq -1, 0$. It is easy to check, that with this choice for h_k and g_k, equations (22.21) and (22.22) will become identical to equations (22.12).

22.9 Non-orthogonal wavelets

Often, the condition of orthogonality of wavelets is is abandoned, but usually equations similar to (22.21) (22.22) are still obtained to perform efficient calculation. Usually, non-orthogonal wavelets

are also linearly dependent. One of the advantages of this is redundant representation of the coefficients $c_{0,k}$ for error correction. For calculations, there is little need to get involved with theoretical issues, and it is enough to know the high-pass and low-pass filters used in calculations,[22.22] but one may need a somewhat closer understanding in order to see how these wavelets can be used.

22.10   Applications of wavelets in finance

Frequency analysis has long been established in engineering and the sciences, but it has major limitations in that it imposes major restrictions on the dynamics on the time series such as stationarity, and information in the time domain is lost. While stationary time series are common in engineering, they are rare in finance.

Wavelets overcome these limitations, and they are capable of capturing information both in the frequency domain and the time domain. The early development of wavelets took place in image analysis. In the last two decades, their applications became wide-spread in the sciences, but they were slow to emerge in finance. This situation is now changing; the paper [22] gives a simple introductions to wavelets, and discusses their applications in finance. The paper argues that the advantages of wavelet methods are that they combine time-domain and frequency-domain information, and, further, that they are very flexible, and do not make strong assumptions about the data generating the time series under consideration. The paper [7] gives a tutorial of the wavelet transform. The doctoral dissertation [28] uses wavelets for financial time series to discuss the interaction between major equity markets, and discusses wavelet networks, a special class of neural networks, in financial forecasting. The master's thesis [32] analyzes various financial model experiments, and demonstrates that wavelet neural networks combined with statistical methods is feasible for achieving accurate forecasting. The paper [31] uses wavelets to analyze the effects of high-frequency trading on the stock market.

23   State-space models

23.1   A simple state-space model

Given a field F (in these notes mainly the field $\mathbb{R}$ of real numbers or $\mathbb{C}$ of complex numbers), write $F_{m,n}$ for the set (or algebra)[23.1] over F of $m \times n$ matrices. A state-space model involves two vector time series: $\{S_t\}$, the state of the system, and $\{Y_t\}$, the observed time series; here for given positive integers m and n we have $S_t \in \mathbb{R}_{n,1}$ and $Y_t \in \mathbb{R}_{m,1}$ are column vectors.[23.2] S_t is not assumed to be known. The updating equations are

$$(23.1) \qquad\qquad\qquad\qquad S_t = AS_{t-1} + e_t,$$
$$(23.2) \qquad\qquad\qquad\qquad Y_t = HS_t + \epsilon_t.$$

Here $e_t \in \mathbf{R}_{n,1}$ is the column vector of errors in the update equation (23.1), $\epsilon_t \in \mathbb{R}_{m,1}$ is the vector of error in the observation equation (23.2). Further, $A \in \mathbf{R}_{n,n}$, and $H \in \mathbf{R}_{m,n}$ are matrices.

Sometimes one also assumes that the e_t are identically distributed; similarly, one may assume that the errors of ϵ_t are identically distributed. Further, one often also assumes that any collection of the vectors e_t and ϵ_t for various values of t is independent. [17]

[22.22]Some wavelet schemes may involve several high-pass and low-pass filters.

[23.1]An algebra over a field F is a vector space over F also has a product operation with certain properties. In a matrix algebra, the product operation is matrix multiplication.

[23.2]What we did is somewhat of an abuse of notation. Namely, the entries of the vector S_t are random variables and not numbers; so saying that $S_t \in \mathbf{R}_{n,1}$ is technically incorrect. Similarly for Y_t.

23.2 Representation of simple state-space models as ARMA models

Let m, n, M, and N be positive integers, Let $\{Y_t\}$ be vector time series, $Y_t \in \mathbb{R}_{m,1}$, let $E_t \in \mathbb{R}_{n,1}$ be identically distributed error vectors such that any collection of them for different values of t is independent. Let $A_k \in \mathbb{R}_{m,m}$ $(1 \leq k \leq M)$ and $B_l \in \mathbb{R}_{m,n}$ $(0 \leq l \leq N)$ be matrices. Assume

$$(23.3) \qquad Y_t = \sum_{k=1}^{M} A_k Y_{t-k} + \sum_{l=0}^{N} B_l E_{t-l}$$

for all t. This equation is called a vector ARMA(M, N) model of the time series $\{Y_t\}$. If the matrices A_k are scalar multiples of the identity matrix, we call the model a vector ARMA(M, N) model with scalar AR coefficients. We have

Theorem 23.1. *The vector Y_t in equations (23.1) and (23.2) satisfies an ARMA(N,N) model with scalar AR coefficients for some $N \leq m$, where m is the dimension of the state vector S_t.*

If the errors e_t in (23.1) are identically distributed, and also so are the errors ϵ_t in (23.2), then the errors in the obtained ARMA model are also identically distributed. Similarly, if the error vectors $(\epsilon_t^*, e_t^*)^*$ are independent,[23.3] then so are the errors in the obtained ARMA model.

Proof. By repeated applications of equation (23.1), we can see that for any integer $k \geq 0$ we have

$$(23.4) \qquad S_{t+k} = A^k S_t + \sum_{j=0}^{k-1} A^j e_{t+k-j}.$$

This is easy to verify by induction. Indeed, for $k = 0$ this says that $S_t = S_t$. Assuming the equation is true with a certain value of k, by equation (23.1) we have

$$S_{t+k+1} = A S_{t+k} + e_{t+k+1} = A\Big(A^k S_t + \sum_{j=0}^{k-1} A^j e_{t+k-j}\Big) + e_{t+k+1}$$

$$= A^{k+1} S_t + \sum_{j=0}^{k} A^j e_{t+k+1-j},$$

establishing equation (23.4). Let $P(x) = \sum_{k=0}^{N} \alpha_k x^k$ $(\alpha_N = 1)$ be the minimal polynomial of the matrix A.[23.4] Multiplying equation (23.4) by α_k and adding the resulting equations for $0 \leq k \leq N$, we obtain

$$\sum_{k=0}^{N} \alpha_k S_{t+k} = \sum_{k=0}^{N} \alpha_k A^k S_t + \sum_{k=0}^{N} \alpha_k \sum_{j=0}^{k-1} A^j e_{t+k-j}$$

$$= P(A) S_t + \sum_{k=0}^{N} \alpha_k \sum_{l=1}^{k} A^{k-l} e_{t+l};$$

[23.3] The repeated application of the Hermitian transpose cancel out, since for any number or matrix x we have $(x^*)^* = x$. We wrote out the right-hand side to illustrate this. The purpose of this notation is to avoid the use of writing column vectors, which take up more space to print.

[23.4] A *monic* polynomial is a polynomial with leading coefficient 1. The minimal polynomial $P(x)$ of an $n \times n$ matrix A is the the monic polynomial of the smallest degree such that $P(A) = 0$. It is known that the degree of the minimal polynomial of A is $\leq n$. This is because we have $Q(A) = 0$ for the characteristic polynomial $Q(x)$ of A by the Cayley–Hamilton theorem. See [24, Subsections 3.1 and 8.8, and especially Theorem 6.1 on p. 14 in Section 6].

in the last equation we replaced the summation variable j with $l = k - j$.[23.5] Noting that $P(A) = 0$, and interchanging the order of summation on the right, we obtain

$$\sum_{k=0}^{N} \alpha_k S_{t+k} = \sum_{l=1}^{N} \sum_{k=l}^{N} \alpha_k A^{k-l} e_{t+l}.$$

Multiplying this equation by the matrix H on the left and using equation (23.2) with $t + k$ replacing t, we obtain that

$$\sum_{k=0}^{N} \alpha_k (Y_{t+k} - \epsilon_{t+k}) = \sum_{l=1}^{N} \sum_{k=l}^{N} \alpha_k H A^{k-l} e_{t+l},$$

Since $\alpha_N = 1$, this equation can be written as

$$Y_{t+N} = -\sum_{k=0}^{N-1} \alpha_k Y_{t+k} + \sum_{l=0}^{N} \alpha_l \epsilon_{t+l} + \sum_{l=1}^{N} \sum_{k=l}^{N} \left(\alpha_k H A^{k-l} \right) e_{t+l}$$

$$= -\sum_{k=0}^{N-1} \alpha_k Y_{t+k} + \sum_{j=0}^{N} \alpha_{N-j} \epsilon_{t+N-j} + \sum_{j=0}^{N-1} \sum_{k=N-j}^{N} \left(\alpha_k H A^{k+j-N} \right) e_{t+N-j},$$

where, to obtain the last equation, we put $j = N - l$ in the summation. To make this equation fit the form described in equation (23.3), take

$$E_t = (\epsilon_t^*, e_t^*)^* = \begin{pmatrix} \epsilon_t \\ e_t \end{pmatrix},$$

where we used Hermitian transpose in the middle member to save space. The matrix on the right is an $m + n$ dimensional column vectors (since ϵ_t is m-dimensional and e_t is n-dimensional). Further,

$$B_j = \left(\alpha_{N-j} I_m, \ \sum_{k=N-j}^{N} \alpha_k H A^{k+j-N} \right) \qquad \text{for} \quad 0 \le j \le N-1,$$

$$B_N = (\alpha_0 I_m, 0_{m,n}) \qquad\qquad\qquad \text{for} \quad j = N,$$

where I_m is the $m \times m$ identity matrix and $0_{m,n}$ is the $m \times n$ zero matrix; since H is $m \times n$ and A is $n \times n$ matrix this makes B_j an $n \times (m + n)$ matrix. As for the comment about the independence and identical distribution of the errors, this is clear from the equations for the error E_t. The proof is complete. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ □

23.3   Representation of an ARMA model as a state-space model

Conversely, an $\text{ARMA}(M, N)$ model can also be represented as a state-space model described by equations (23.1) and (23.2). We will only consider a scalar ARMA model of form

$$Y_t = \sum_{k=1}^{M} \phi_k Y_{t-k} + \sum_{l=0}^{N} \theta_l v_{t-l},$$

[23.5]Even though A^k for negative k occurs in the above equations, and A^{-1} may not be defined, this is harmless, since the coefficient of A^k for negative k is 0.

where $\{Y_t\}$ is a scalar time series, and the errors v_t are scalar. To represent this as a state-space model, we take the state as the vector

$$S_t = (Y_t^*, Y_{t-1}^*, \ldots, Y_{t-M+1}^*)^*,$$

where the repeated application of the Hermitian transpose cancels out – see footnote 23.3 on p. 92. Take $A = (a_{ij})$ be an $M \times M$ matrix with $a_{1j} = \phi_k$ for $1 \le j \le M$, $a_{ij} = \delta_{i-1,j}$ for $2 \le i \le M$ and $1 \le j \le M$, and $e_t = 0_{M,1}$. Further, let H be the N-dimensional row vector

$$H = (1, 0, 0, \ldots, 0),$$

and let ϵ_t be the scalar

$$\epsilon_t = \sum_{l=0}^{N} \theta_l v_{t-l}.$$

With these choices, equations (23.1) and (23.2) are satisfied. As it is seen from these equations, the independence of the errors ϵ_t is not assured in this model.

23.4 Question whether the ARMA model of scalar time series with scalar errors is appropriate

It seems that in a scalar ARMA model the presence of past errors is an artifact. It is reasonable to assume that the past behavior of a system producing a time series is communicated via the current state of a system. That is, the correct model of a time series would apparently be a state-space model, and the ARMA behavior is only a mathematical consequence of the state-space model. However, as we saw in Theorem 23.1, the errors in this ARMA model have matrix coefficients, and an ARMA model with scalar error coefficients could produce only a relatively poor approximation of the actual errors. For this reason, the description showing how past errors in the ARMA model seems more of a mathematical artifact than some philosophical reflection on the behavior a system.

The translation of an ARMA model into a state-space model is only a mathematical trick in that we describe the state as a vector of past outputs, and is not based on a deeper understanding of the system producing the signals. The fact that independent errors in the time series are not reflected in the independence of the state errors at different times of the state-space model obtained from the ARMA model points even more to the artificialness of the ARMA model.

24 The Kalman filter

24.1 What is the Kalman filter trying to do?

Imagine you are steering on ship through a narrow and dangerous straight. You can control the steering and the engine power. However, the ship is slow to respond to any input. Furthermore, the ship's response has a random element because of currents and wind. You have precise maps indicating the route the ship is required to follow. You can monitor the ship's position and orientation (perhaps by GPS and compass, or by features on land visible from the ship). The position of the ship, her orientation, the position of the steering wheel, the engine controls, and the actual engine power is monitored as the sampled values vector time series. It is also known how the ship is supposed to respond to steering and engine controls. The problem to be solved is how to change the steering and engine controls to keep the ship safe. The ship's response to these controls involves various delays and random elements, so the exact state of the engine is not known; all information about

it comes from various sensors. A mathematical method to handle this situation was invented by Rudolf E. Kalman. The mathematical model encompassing his method will be described next.

24.2 A state-space model with control input

A time series model of the described situation can in general be described as follows; we will assume in the rest of this section that all matrices are real. We are trying to estimate the state vector $X_k \in \mathbb{R}_{n \times 1}$ (i.e., an n-dimensional column vector of reals) of a process at time k with measurements $Z_k \in R_{m \times 1}$, where m and n are positive integers. Here Z_k is known but X_k is not. These quantities are governed by the following equations

$$(24.1) \qquad\qquad\qquad X_k = F_k X_{k-1} + B_k u_k + e_k,$$

$$(24.2) \qquad\qquad\qquad Z_k = H_k X_k + \epsilon_k.$$

Here $u_k \in \mathbb{R}_{p,1}$ is the control input at time k, where p is an integer, $F_k \in \mathbb{R}_{n,n}$ is the state transition matrix applied to the previous state, $B_k \in \mathbb{R}_{n,p}$ is the control-input model that is applied to the control input u_k, $H_k \in \mathbb{R}_{m,n}$ is the observation model, $e_k \in \mathbb{R}_{n,1}$ is the control error, and $\epsilon_k \in \mathbb{R}_{m,1}$ is the measurement error, all at time k. The errors e_k and ϵ_k are assumed to follow multivariate normal distribution with 0 mean with covariance matrices $Q_k \in \mathbb{R}_{n,n}$ and $R_k \in \mathbb{R}_{p,p}$.[24.1] We will assume that the errors e_k and ϵ_k are independent of each other and of any X_l, Z_l e_l, and ϵ_l for $l < k$. We further assume that ϵ_k is independent of X_k. It is not necessary to know the covariance matrices Q_k and R_k; they can be estimated from prior observations (called the tuning of the process). In equations (24.1) and (24.2), the matrices F_k, B_k, and H_k are assumed to be known.[24.2]

24.3 The Kalman filter: prediction

The Kalman filter works in two steps: a prediction step, and an update step. $\hat{X}_{k|k-1}$ denotes the predicted estimate of X_k before the measurement Z_k is taken into account, and $\hat{X}_{k|k}$ denotes the corrected estimate after the measurement Z_k is known. We will assume that these estimates are unbiased, that is, their mean is $E(X_k)$.[24.3] Note that X_k is not observable. We put

$$(24.3) \qquad\qquad\qquad \hat{X}_{k|k-1} = F_k \hat{X}_{k-1|k-1} + B_k u_k.$$

We have

$$(24.4) \qquad\qquad\qquad X_k - \hat{X}_{k|k-1} = F_k(X_{k-1} - \hat{X}_{k-1|k-1}) + e_k$$

according to (24.1). Since $E(e_k) = 0$, it follows that if $\hat{X}_{k-1|k-1}$ is an unbiased estimator of X_{k-1}, then $\hat{X}_{k|k-1}$ is an unbiased estimator of X_k. We will consider the covariance matrices P of the errors of these estimators. That is,

$$(24.5) \qquad\quad P_{k|k} = \text{Cov}(X_k - \hat{X}_{k|k}) = \text{E}\big((X_k - \hat{X}_{k|k})(X_k - \hat{X}_{k|k})^T\big),$$

$$(24.6) \qquad\quad P_{k|k-1} = \text{Cov}(X_k - \hat{X}_{k|k-1}) = \text{E}\big((X_k - \hat{X}_{k|k-1})(X_k - \hat{X}_{k|k-1})^T\big);$$

[24.1]In symbols, one can write that $e_k \sim \mathcal{N}(0_{n,1}, Q_k)$ and $\epsilon_k \sim \mathcal{N}(0_{p,1}, R_k)$, where, given positive integers l and m, $0_{l,m} \in \mathbb{R}_{l,m}$ denotes an $l \times m$ matrix with all zero entries. We will also write $Q_k = \text{Cov}(e_k)$ and $R_k = \text{Cov}(\epsilon_k)$. This notation for the covariance matrix was introduced in Subsection 2.3.

[24.2]The model we are describing is a linear model. The extended Kalman filter is a nonlinear model, in which these matrices are Jacobian matrices of the variables at places of the variables X and Z known or estimated at time k. See [34, p. 8].

[24.3]This will be asymptotically true if the filter converges. When the filter is started, the value of $X_{0|0}$ will be a guess.

The second equation on each line assumes that these estimators are unbiased. Using equations (24.4) and (2.5), and noting that $E(e_k) = 0$ and $\text{Cov}(e_k) = Q_k$, we obtain

$$
\begin{aligned}
(24.7) \qquad P_{k|k-1} &= \text{E}\Big(\big(F_k(X_{k-1} - \hat{X}_{k-1|k-1}) + e_k\big)\big((X_{k-1} - \hat{X}_{k-1|k-1})^T F_k^T + e_k^T\big)\Big) \\
&= F_k \, \text{E}\big((X_{k-1} - \hat{X}_{k-1|k-1})(X_{k-1} - \hat{X}_{k-1|k-1})^T\big) F_k^T + \text{E}(e_k e_k^T) \\
&= F_k P_{k-1|k-1} F_k^T + Q_k;
\end{aligned}
$$

the second equation here holds because the error vector e_k is independent of earlier variables. The third equation uses (24.5) and the equation $Q_k = \text{Cov}(e_k)$.

The measurement residual is given by

$$
(24.8) \qquad \tilde{Y}_k \stackrel{def}{=} Z_k - H_k \hat{X}_{k|k-1}.
$$

According to equation (24.2), we would have $\tilde{Y}_k = 0$ if $\hat{X}_{k|k-1}$ were accurate. That is, $\tilde{Y}_k$ is a measure of the accuracy of the prediction. Note that

$$
\text{E}(\tilde{Y}_k) = \text{E}(Z_k) - H_k \, \text{E}(\hat{X}_{k|k-1}) = H_k \, \text{E}(X_k) - H_k \, \text{E}(\hat{X}_{k|k-1}) = 0,
$$

where the second equation holds in view of (24.2); for the third equation, see (24.4), where we remarked that $\hat{X}_{k|k-1}$ is an unbiased estimator of X_k. Hence, using equation (24.2) once more, for the covariance matrix of $\tilde{Y}_k$ we have

$$
\begin{aligned}
(24.9) \qquad S_k \stackrel{def}{=} \text{Cov}(\tilde{Y}_k) &= \text{E}(\tilde{Y}_k \tilde{Y}_k^T) = \text{E}\Big(\big(H_k(X_k - \hat{X}_{k|k-1}) + \epsilon_k\big)\big((X_k - \hat{X}_{k|k-1})^T H_k^T + \epsilon_k^T\big)\Big) \\
&= H_k \, \text{E}\big((X_k - \hat{X}_{k|k-1})((X_k - \hat{X}_{k|k-1})^T)\big) H_k^T + \text{E}(\epsilon_k \epsilon_k^T) \\
&= H_k P_{k|k-1} H_k^T + R_k \, ;
\end{aligned}
$$

the second equation here holds because ϵ_k is independent of X_k and $\hat{X}_{k|k-1}$, and $\text{E}(\epsilon_k) = 0$; finally, the third equation holds in view of (24.6) and since $R_k = \text{Cov}(\epsilon_k)$.

Before continuing, some reflection can be helpful. While the quantities X_k, Z_k, and $\tilde{Y}_k$, and the estimators $\hat{X}_{k|k-1}$, X_k, and $\hat{X}_{k|k}$ are random variables, these quantities are functions of the sample space. On the other hand, the matrices $P_{k|k}$, $P_{k|k-1}$, S_k, and others are not. They are not functions on the sample space, they are functions only of the expectations of various expressions of the random variables involved. On the other hand, in an implementation of the Kalman filter, these expectations may be approximated by random variables.

24.4   The Kalman filter: the correction

The residual $\tilde{Y}_k$ carries the information about the accuracy of the estimate $\hat{X}_{k|k-1}$. We define the corrected estimate

$$
(24.10) \qquad \hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k \tilde{Y}_k,
$$

where the matrix K_k is so chosen that the mean square error

$$
(24.11) \qquad \text{E}(\|X_k - \hat{X}_{k|k}\|^2)
$$

is the least possible. Noting that

$$\mathrm{E}(X_k - \hat{X}_{k|k}) = \mathrm{E}(X_k - \hat{X}_{k|k-1}) - K_k \,\mathrm{E}(\tilde{Y}_k) = 0,$$

we have

(24.12) $$P_{k|k} \stackrel{def}{=} \mathrm{Cov}(X_k - \hat{X}_{k|k}) = \mathrm{E}\big((X_k - \hat{X}_{k|k})(X_k - \hat{X}_{k|k})^T\big);$$

this equation is in fact a restatement of equation (24.5), first stated in anticipation of the definition of $\hat{X}_{k|k}$. The *trace*[24.4] of this matrix is $\mathrm{E}(\|X_k - \hat{X}_k\|^2)$. That is K_k is to be determined so as to minimize the trace of $P_{k|k}$. Using equations (24.10), (24.9), and (24.6), we have

(24.13)
$$
\begin{aligned}
P_{k|k} &= \mathrm{E}\Big(\big((X_k - \hat{X}_{k|k-1}) - K_k\tilde{Y}_k\big)\big((X_k - \hat{X}_{k|k-1})^T - \tilde{Y}_k^T K_k^T\big)\Big) \\
&= \mathrm{E}\big(X_k - \hat{X}_{k|k-1})\big)(X_k - \hat{X}_{k|k-1})^T\big) + K_k\,\mathrm{E}(\tilde{Y}_k\tilde{Y}_k^T)K_k^T \\
&\quad - \mathrm{E}\big((X_k - \hat{X}_{k|k-1})\tilde{Y}_k^T\big)K_k^T - K_k\,\mathrm{E}\big(\tilde{Y}_k(X_k - \hat{X}_{k|k-1})^T\big) \\
&= P_{k|k-1} + K_k S_k K_k^T - \mathrm{E}\big((X_k - \hat{X}_{k|k-1})\tilde{Y}_k^T\big)K_k^T - K_k\,\mathrm{E}\big(\tilde{Y}_k(X_k - \hat{X}_{k|k-1})^T\big).
\end{aligned}
$$

We need to simplify the last two terms. We first deal with the last term. Using equations (24.8) and (24.2), we obtain

$$
\begin{aligned}
\mathrm{E}\big(\tilde{Y}_k(X_k - \hat{X}_{k|k-1})^T\big) &= \mathrm{E}\big((Z_k - H_k\hat{X}_{k|k-1})(X_k - \hat{X}_{k|k-1})^T\big) \\
&= \mathrm{E}\big((H_k X_k + \epsilon_k - H_k\hat{X}_{k|k-1})(X_k - \hat{X}_{k|k-1})^T\big) \\
&= H_k\,\mathrm{E}\big((X_k - \hat{X}_{k|k-1})(X_k - \hat{X}_{k|k-1})^T\big) = H_k P_{k|k-1}\,;
\end{aligned}
$$

here, the penultimate[24.5] equation follows since ϵ_k is independent of X_k and $\hat{X}_{k|k-1}$, and $\mathrm{E}(\epsilon_k) = 0$; finally, the the last equation holds in view of (24.6). Observing that the second term on the right of (24.13) is just the transpose of the third term, using this (24.13) becomes

(24.14) $$P_{k|k} = P_{k|k-1} + K_k S_k K_k^T - P_{k|k-1} H_k^T K_k^T - K_k H_k P_{k|k-1}\,;$$

for the last term note that $P_{k|k-1}$, being a covariance matrix, is symmetric, so $P_{k|k-1}^T = P_{k|k-1}$.

24.5   Optimization of the Kalman gain

In equation (24.14) all the matrices are known at this point except for the matrix K_k. To determine the optimal gain, we need to choose the matrix such that the trace of $P_{k|k}$ is the smallest possible. This problem is always solvable, since the trace of this matrix is a positive semi-definite quadratic form, with the entries of K_k being the variables. Indeed, this trace is the expression given in (24.11). This problem can be solved as a simple problem of optimization in multivariate calculus; however, to avoid technical complications, we need the right mathematical symbolism. There are several mathematical approaches that could be used: we could write out the trace in question with sums of products involving scalar variables and then take partial derivatives; to simplify the calculations, we could use matrix differential calculus (see [38]), or we could use tensor calculus. We wish to avoid these complications, since the same goal can be accomplished making an informal use of infinitesimal matrices.

[24.4]The trace of a square matrix is the sum of its diagonal elements.
[24.5]The one before the last.

The term *infinitesimal* was introduced by Leibniz, and they formed the basis of Leibniz's development of calculus. They denote numbers very close to zero; sometimes in a contradictory way a positive infinitesimal is described as a positive number that is smaller than every "usual" positive real numbers. There are various orders of infinitesimals: if x and y are both infinitesimal and x/y is also an infinitesimal, then x is said to be an infinitesimal of higher order than y. Leibniz's idea was very fruitful, and they led to a fast development of calculus. In the 19th century, infinitesimals were exiled from mathematics, and replaced by "precise" mathematical tools.

We put "precise" in quotes, since there are different levels of precision acknowledged by mathematical logic. In fact, Kurt Gödel in the 20th century showed the limitations of formal approaches to mathematics; meanwhile, Gödel's ideas via Alan Turing and John von Neumann led the way to modern computer architecture (see [26]) The ideas of Skolem and Gödel led to models satisfying the axioms of arithmetic that different from the usual (standard) set of integers. Considerations of such models inspired Abraham Robinson to invent nonstandard analysis, and which put infinitesimals on a rigorous mathematical foundation; his book [29] is still the best source the learn the subject from. He and Allen R. Bernstein did nonstandard analysis to good use, and in 1966 solved a problem involving invariant subspaces of Hilbert spaces. Perhaps to the misfortune of nonstandard analysis,[24.6] but very much to the fortune of mathematics, in 1973, V. I. Lomonosov of the Soviet Union, who later emigrated to the USA, gave a striking generalization of the Bernstein–Robinson result – see [37]. The Wikipedia article [40] is a good overview of the subject. If you happen to look also at the article Criticism of non-standard analysis also on Wikipedia [39], the criticism is somewhat misguided. Its main role is not to establish a philosophical basis for infinitesimals; it is a mathematical tool to simplify a number of argument, somewhat similar in the way general topology is such a tool.

In finding the optimal choice of K_k in equation (24.14), we replace K_k with $K_k + hM$, where h is an infinitesimal scalar, and M is an arbitrary matrix, and we write the resulting matrix on the left-hand side as $P_{k|k} + \Delta P_{k|k}$; that is, $\Delta P_{k|k}$ represents the change in the matrix $P_{k|k}$ by this replacement:

$$
\begin{aligned}
P_{k|k} &+ \Delta P_{k|k} \\
&= P_{k|k-1} + (K_k + hM)S_k(K_k^T + hM^T) - P_{k|k-1}H_k^T(K_k^T + hM^T) - (hM + K_k)H_kP_{k|k-1} \\
&= P_{k|k-1} + K_kS_kK_k^T - P_{k|k-1}H_k^TK_k^T - K_kH_kP_{k|k-1} \\
&\quad + h(MS_kK_k^T + K_kS_kM^T - P_{k|k-1}H_k^TM^T - MH_kP_{k|k-1}) + h^2MS_kM^T \\
&= P_{k|k} + h(MS_kK_k^T + K_kS_kM^T - P_{k|k-1}H_k^TM^T - MH_kP_{k|k-1}) + h^2MS_kM^T,
\end{aligned}
$$

where the last equation holds in view of (24.14). That is,

$$
\Delta P_{k|k} = h(MS_kK_k^T + K_kS_kM^T - P_{k|k-1}H_k^TM^T - MH_kP_{k|k-1}) + h^2MS_kM^T.
$$

Denoting by $\delta P_{k|k}$ what remains of $\Delta P_{k|k}$ after omitting the higher order infinitesimals, i.e., term multiplied by h^2, we have $\delta P_{k|k} = hD$,[24.7] where

$$
D = MS_kK_k^T + K_kS_kM^T - P_{k|k-1}H_k^TM^T - MH_kP_{k|k-1}.
$$

We are only interested in the trace of this matrix, since we want to minimize the trace of $P_{k|k}$. Denote by $\mathrm{Tr}(A)$ of a matrix A, and note that $\mathrm{Tr}(A) = \mathrm{Tr}(A^T)$, since taking transpose does not change the diagonal elements of a matrix. Observing that in the expression on the right-hand side

---

[24.6]Not really. Nonstandard analysis is well and alive.

[24.7]At the price of some minor additional circumlocution, the matrix D could be described as the directional derivative of $P_{k|k}$ with respect to K_k in the direction of M.

of the equation for D the second term is the transpose of the first, and the third is the transpose of the fourth, we can write that

$$\text{Tr}(D) = 2\,\text{Tr}(MS_k K_k^T - MH_k P_{k|k-1}) = 2\,\text{Tr}\big(M(S_k K_k^T - H_k P_{k|k-1})\big).$$

We want to make $\text{Tr}(D) = 0$, for all M. To this end it is sufficient to to make sure that

$$S_k K_k^T - H_k P_{k|k-1} = 0.$$

In fact, this condition is also necessary, but we will not make use of this.[24.8] Taking transpose, this means that

(24.15) $$\qquad\qquad\qquad K_k S_k - P_{k|k-1} H_k^T = 0\,;$$

note that $P_{k|k-1} = P_{k|k-1}^T$ and $S_k = S_k^T$, since they are covariance matrices (cf. (24.6) and (24.9)), and covariance matrices are symmetric. So, if S_k is invertible, we take

(24.16) $$\qquad\qquad\qquad K_k = P_{k|k-1} H_k^T S_k^{-1}.$$

If S_k is singular, then equation (24.15) is not be solvable or it has multiple solutions, and one needs to make alternative arrangements. Perhaps one would take $K_k = 0$, or else take a subspace, such as that formed by the maximum number of linearly independent columns of S_k, and ensure that equation (24.15) restricted to this subspace is satisfied. Probably the former choice is simpler, since in the next update step it is unlikely that the matrix S_{k+1} is again singular.

24.6 Summary of the Kalman filter steps

We summarize here how these equations are used to operate the Kalman Filter. We start with initial values $\hat{X}_{0|0}$ and $P_{0|0}$. At the k step, we do the updating in two stages. The prediction stage performs those calculations that can be performed before the measurement Z_k comes in. The calculations in the correction stage rely on the measurement Z_k.

The equations for the prediction stage rely on equations (24.3), (24.7), (24.9), (24.16), and (24.14) in turn. They are

$$\hat{X}_{k|k-1} = F_k \hat{X}_{k-1|k-1} + B_k u_k,$$
$$P_{k|k-1} = F_k P_{k-1|k-1} F_k^T + Q_k,$$
$$S_k = H_k P_{k|k-1} H_k^T + R_k,$$
$$K_k = P_{k|k-1} H_k^T S_k^{-1},$$
$$P_{k|k} = P_{k|k-1} + K_k S_k K_k^T - P_{k|k-1} H_k^T K_k^T - K_k H_k P_{k|k-1}.$$

[24.8]The necessity of this condition can be seen as follows. If for a matrix M, the matrix MA has a nonzero element, then pick one of the nonzero elements of MA, and change all elements of M to zero except those that are in the same row as the element picked. This will make MA have a nonzero element only in the row with the element picked. Then, for a permutation matrix P, i.e., a matrix that has exactly one 1 in each row and each column, and all other entries are 0, the rows of PMA are a permutation of the rows of MA. By taking an appropriate permutation matrix, a nonzero element of MA can be moved to the main diagonal. This matrix will have exactly one nonzero element in its main diagonal, so its trace will not be zero.

Hence, if the trace of MA is zero for every matrix M (of the appropriate size), then A has to be the zero matrix.

The equations for the correction stage rely on equations (24.8) and (24.10), respectively. They are

$$\tilde{Y}_k = Z_k - H_k \hat{X}_{k|k-1},$$
$$\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k \tilde{Y}_k.$$

The matrices F_k, B_k, Q_k, and R_k in these equations are assumed to be known in advance, and are not part of the update process.

In a practical installation of the filter, Q_k and R_k can be estimated by the filter itself. Starting with initial estimates Q_0 and R_0 that may be little more than a guess, one estimates e_k from equation (24.1), replacing X_{k-1} and X_k with $\hat{X}_{k-1|k-1}$ and $\hat{X}_{k|k}$. There is no better choice, since the values of X_{k-1} and X_k cannot be known. Similarly, ϵ_k is estimated from equation (24.2), replacing X_k with $\hat{X}_{k|k}$. The estimates for the covariance matrix Q_k and R_k are then updated, using the estimates for these error sequences. This process is called the tuning (the parameters of) the Kalman filter. This tuning of the Kalman filter may itself involve sophisticated algorithms.

25 The extended Kalman filter

25.1 Fréchet derivative

Definition 25.1. Let V be a normed vector space over $\mathbb{R}$. A subset B of V is called an *open ball* if $B = \{x \in V : \|x - c\| < \rho\}$ for some $c \in V$ and for some $\rho > 0$; c is called the center of B and ρ, its radius. A set $S \subset V$ is called open if for every $x \in S$ there is an open ball B with center x such that $B \subset S$.

Definition 25.2. Let V and W be vector spaces over $\mathbb{R}$, and $\Lambda : V \to W$ be a mapping. Λ is called a *linear operator* (or a *linear transformation*) if $\Lambda(\alpha x + y) = \alpha \Lambda(x) + \Lambda(y)$ for every $x, y \in V$ and every $\alpha \in \mathbb{R}$.

For a linear operator Λ and a vector x, one often writes Λx instead of $\Lambda(x)$.

Definition 25.3. Let V and W be vector spaces over $\mathbb{R}$, and and let $\Lambda : V \to W$ be a linear operator. Λ is called *bounded* if there is an $\alpha \in \mathbb{R}$ such that $\|\Lambda x\|_W \leq \alpha \|x\|_V$ for all $x \in V$, where $\|\cdot\|_V$ and $\|\cdot\|_W$ indicate the norms of the respective spaces. The least such α is called the *norm* of Λ, or, more precisely, its *norm induced* by the vector norms in V and W.

It is easy to see that if an Λ is bounded then there exists a least such α, and, in fact,

$$\|\Lambda\| = \sup\{\|\Lambda x\|_W : x \in V \quad \text{and} \quad \|x\|_V = 1\}.$$

Next, we will describe what is meant by the limit of a function.

Definition 25.4. Let V and W be normed vector spaces over $\mathbb{R}$, let S be a subset of V, let $f : S \to W$ be a function. let $x \in V$, let y run over elements of V, and let $w \in W$. We say that

$$\lim_{y \to x,\ y \in S} f(y) = w$$

if for every $\epsilon > 0$ there is a $\delta > 0$ such that we have $\|f(y) - w\|_W < \epsilon$ whenever $0 < \|y - x\|_V < \delta$ and $y \in S$; here $\|\cdot\|_V$ and $\|\cdot\|_W$ indicate the norms of the respective spaces.

In the definition, saying that $0 < \|y - x\|$ is just another way of saying that $y \neq x$, but it is more concise to the inequality $0 < \|y - x\|_V < \delta$ instead of saying that $\|y - x\|_V < \delta$ and $y \neq x$. In case S is an open set and $x \in S$, we usually write $\lim_{y \to x} f(y)$ instead of $\lim_{y \to x, \, y \in S} f(y)$, since in this case $\|y - x\|_V < \delta$ implies $y \in S$ for small enough δ. The above is the Cauchy definition of limit. which is well known to be equivalent to the Heine definition, according to which

$$\lim_{y \to x, \, y \in S} f(y) = w$$

if for any sequence $\{y_n\}$ of elements of S such that

$$\lim_{n \to \infty} \|y_n - x\| = 0$$

we have

$$\lim_{n \to \infty} \|f(y_n) - w\| = 0.$$

Definition 25.5. Let V and W be normed vector spaces over $\mathbb{R}$, let S be a subset of V, let $f : S \to W$ be a function. let $x \in S$. We say that f is *continuous at x in S* if $\lim_{y \to x, \, y \in S} f(y) = f(x)$. If there is an open ball $B \subset S$ with center x, then we simply say that f is continuous at x.

Definition 25.6. Let V and W be normed vector spaces over $\mathbb{R}$, let $S \subset V$ be open, let $f : S \to W$ be a function. let $x \in S$, and let $\Lambda : V \to W$ be a bounded linear operator. Λ is called the Fréchet derivative of f at $x \in V$ if, with y running over elements of V, we have

$$\lim_{y \to x} \frac{\|f(y) - f(x) - \Lambda(y - x)\|_W}{\|y - x\|_V} = 0.$$

25.2   The Jacobian matrix

Let V be an n-dimensional real euclidean space. That is, $V = \mathbb{R}_{n,1}$ is the space of $n \times 1$ column vectors with real entries, and for $x = (\xi_1, \xi_2, \ldots, \xi_n)^T \in V$, the norm of V is defined as

$$\|x\|_V = \sqrt{x^T x} = \left(\sum_{k=1}^{n} \xi_k^2 \right)^{1/2}.$$

If $y = (\eta_1, \eta_2, \ldots \eta_n)^T$ is another element of V, then the euclidean inner product is defined as

$$\langle x, y \rangle_V = x^T y = \sum_{k=1}^{n} \xi_k \eta_k.$$

Since the $\|\cdot\|_V$ is the norm induced by the inner product $\langle \cdot, \cdot \rangle_V$, it follows that $\|\cdot\|_V$ is indeed a norm; in particular, it satisfies Minkowski's inequality; see Clause (c) in Definition 17.2 and Problem 17.2.

Given an m-dimensional real euclidean space W, a function $f : V \to W$ can be described by m functions of n variables: if $f(x) = w$ for $x = (\xi_1, \xi_2, \ldots, \xi_n)^T \in V$, and $w = (\omega_1, \omega_2, \ldots, \omega_m)^T \in W$, writing $\omega_l = f_l(\xi_1, \xi_2, \ldots, \xi_n)$ for l with $1 \leq l \leq m$, these functions describe the function f. In a shortened notation, we may write that $\omega_l = f_l(x)$.

The matrix described in the next definition is called the *Jacobian matrix*, named after the German mathematician Carl Gustav Jacob Jacobi.

Definition 25.7. If V, W, f, f_l, x, and w are as described, the $m \times n$ matrix

(25.1) $$\frac{\partial w}{\partial x} = \frac{\partial f(x)}{\partial x} = \frac{\partial(\omega_1, \omega_2, \ldots, \omega_m)}{\partial(\xi_1, \xi_2, \ldots, \xi_n)} \overset{def}{=} \left(\frac{\partial f_l(\xi_1, \xi_2, \ldots, \xi_n)}{\partial \xi_k} \right)_{1 \le l \le m,\ 1 \le k \le n}$$

is called the *Jacobian matrix* of f at $x = (\xi_1, \xi_2, \ldots, \xi_n)^T$, assuming that the partial derivatives exist.

There is a problem with the notation used for partial derivatives in describing the Jacobian in equation (25.1). To introduce a better notation, we will write ∂_k for the the partial derivative with respect to the kth variable. With this notation, we will write

$$\frac{\partial f_l(x)}{\partial \xi_k} = \partial_k f_l(x) \qquad (1 \le l \le m)$$

The problem with the notation used on the left-hand side is that it is associated with the point $x = (\xi_1, \xi_2, \ldots, \xi_n)^T \in V$. For the point $t = (\tau_1, \tau_2, \ldots, \tau_n)^T \in V$, we would have to write $\partial_k f_l(t) = \partial f_l(t)/\partial \tau_k$. Along the same line, for the Jacobian we can write ∂f. That is

$$\partial f(x) \overset{def}{=} \frac{\partial f(x)}{\partial x}.$$

Lemma 25.1. *Let V, W, f, f_l, x, and w be as described. Assume that the partial derivatives $\partial f_l(x)/\partial \xi_k$ are continuous at x. Then the Jacobian $J = \partial f(x)/\partial x$, interpreted as the linear operator $J : V \to W$ with $J(x) = Jx$ (the right-hand side indicating matrix multiplication) is the Fréchet derivative of f at x.*

Proof. For the partial derivatives of f to be continuous at x there must be an open ball with center x in which these partial derivatives exist; let B be such an open ball, and let $y = (\eta_1, \eta_2, \ldots, \eta_n)^T \in B$ be different from x. Let $\delta_k = \eta_k - \xi_k$ for k with $1 \le k \le n$, and let $h_k \in V$ be the vector all whose components are 0 except that its kth component is δ_k. Let $x_k = x + \sum_{j=1}^{k} h_j$ for k with $0 \le k \le n$. Then we have $x = x_0$ and $y = x_n$. Furthermore, for any k and l with $1 \le k \le n$ and $1 \le l \le m$, we have

(25.2) $$f_l(x_k) - f_l(x_{k-1}) = \delta_k \partial_k f_l(x_{k-1} + \theta_{kl} h_k)$$

for some θ_k with $0 < \theta_{kl} < 1$ by the Mean-Value Theorem of Differentiation, as we will explain. First note that, given that $y \in B$, we have $x_{k-1}, x_k \in B$, and so $f_l(x_{k-1} + \theta h_k)$ as a function of θ is differentiable in the interval $[0, 1]$; indeed,

$$\frac{df_l(x_{k-1} + \theta h_k)}{d\theta} = \delta_k \partial_k f_l(x_{k-1} + \theta h_k).$$

so we can use the Mean-Value Theorem.[25.1] Noting that $\partial_k f_l$ is continuous at x, equation (25.2), can be written as

$$f_l(x_k) - f_l(x_{k-1}) = \delta_k \left(\partial_k f_l(x) + \epsilon_{kl}(y) \right),$$

[25.1]The Mean-Value Theorem says that if ϕ is continuous in the interval $[a, b]$ where $a < b$ and is differentiable in (a, b), then there is a $\xi \in (a, b)$ such that

$$\phi(b) - \phi(a) = \phi'(\xi)(b - a).$$

The Mean-Value Theorem is used in case $h_k \ne 0$; the equation is obviously true also in case $h_k = 0$ (note that $h_k = 0$ is allowed, even though $h_k = 0$ cannot be true for every k, since $y \ne x$).

with some $\epsilon_{kl}(y)$ such that

(25.3)
$$\lim_{y \to x} \epsilon_{kl}(y) = 0,$$

where the dependence of $\epsilon_{kl}(y)$ on x is not indicated, since x is fixed throughout this argument; note that on the left-hand side of equation (25.2), x_{k-1} and x_k are determined by y (and x). Hence, we have

$$f_l(y) - f_l(x) = \sum_{k=1}^{n} \big(f_l(x_k) - f_l(x_{k-1})\big) = \sum_{k=1}^{n} \delta_k \big(\partial_k f_l(x) + \epsilon_{kl}(y)\big)$$

$$= \sum_{k=1}^{n} \delta_k \partial_k f_l(x) + \sum_{k=1}^{n} \delta_k \epsilon_{kl}(y) = \big(J(y-x)\big)_l + \sum_{k=1}^{n} \delta_k \epsilon_{kl}(y),$$

where $\big(J(y-x)\big)_l$ denotes the lth component of the vector $J(y-x)$; the last equation follows from the definition of the Jacobian J and by noting that $y - x = (\delta_1, \delta_2, \dots \delta_n)^T$. That is, writing $e_l \in W$ for the vector all whose components are 0 except that its lth component is 1, we have

$$f(y) - f(x) = J(y-x) + \sum_{l=1}^{m} \sum_{k=1}^{n} \delta_k \epsilon_{kl}(y)\, e_l.$$

Noting that

$$\|y - x\|_V = \Big(\sum_{k=1}^{n} \delta_k^2\Big)^{1/2},$$

we have $|\delta_k| \le \|y - x\|_V$. Hence

$$\|f(y) - f(x) - J(y-x)\|_W \le \sum_{l=1}^{m} \sum_{k=1}^{n} \|y - x\|_V \, |\epsilon_{kl}(y)| \, \|e_l\|_W$$

$$= m \sum_{k=1}^{n} \|y - x\|_V \, |\epsilon_{kl}(y)|;$$

the last equation follows since $\|e_l\|_W = 1$. Therefore, (25.3) implies that

$$\lim_{y \to x} \frac{\|f(y) - f(x) - J(y-x)\|_W}{\|y - x\|_V} = 0,$$

so J is indeed the Fréchet derivative of f at x. $\qquad\square$

In case $m = n = 1$, the Jacobian is just the ordinary derivative of f, and the linear approximation to f implied by the Fréchet derivative described by the Jacobian is just the tangent line to the graph of f. In case $m = 1$ and $n = 2$, the Jacobian describes the total differential of f, and the linear approximation implied by the Fréchet derivative is the tangent plane to the surface given by f. The cases $m = 2$ and $n = 2$ or $m = 3$ and $n = 3$ are occasionally discussed in introductory college courses in the context of changing variables in multiple integrals; such an application was discussed in Subsection 2.4 on account of determining the density function of a nondegenerate multivariate normal distribution.

25.3 The extended Kalman filter

In the model for the extended Kalman filter, the linear equations (24.1) and (24.2) are replaced by nonlinear equations

$$(25.4) \qquad\qquad X_k = f(X_{k-1}, u_k, e_k),$$
$$(25.5) \qquad\qquad Z_k = h(X_k, \epsilon_k).$$

for given vector-valued functions f and g. This can be handled by a modification of equations (24.1) and (24.2) of the linear Kalman filter:

$$X_k = F_k X_{k-1} + B_k u_k + E_k e_k,$$
$$Z_k = H_k X_k + G_k \epsilon_k.$$

In these equations, the coefficient matrices E_k and G_k are new as compared to equations (24.1) and (24.2). The coefficient matrices F_k, B_k, E_k, H_k, and G_k are taken to be the Jacobian matrices[25.2] with respect to the variables associated with these coefficient matrices of f and g, at the place $(\hat{X}_{k-1}, u_k, 0)$ for f, and at $(\hat{X}_k, 0)$ for g, where $\hat{X}_{k-1}$ and $\hat{X}_k$ are the estimates for X_{k-1} and X_k, and the actual value of the vector u_k. The best estimate for e_k and ϵ_k is 0, that is why 0 is taken for the arguments representing these errors.

25.4 Applications of the Kalman-filter

The paper [17] describes the example of a train moving on a straight track, illustrating were the matrices F_k, B_k, and H_k can be obtained from equations of physics describing the system. The Kalman filter was used in aiding landing and return of the lunar module of the Apollo 11 mission, the first human landing on the moon. Today, there are several Kalman filters running on a common cell phone. There is a good description, listing several applications, in the Wikipedia article [36]. The paper [8] lists many more applications; it also contains very interesting details of of these applications.

26 The GARCH model

In an ARMA or ARIMA model

$$(26.1) \qquad\qquad Y_t = \sum_{k=1}^{p} \phi_k Y_{t-k} + e_t - \sum_{l=1}^{q} \theta_k e_{l-k},$$

one often assumes that the errors (or residuals, or innovations) e_t are identically distributed, in particular, they have the same standard deviation. This is often not appropriate for financial time series, which often go through periods of volatility. In these cases, one may prefer to model the errors in the form

$$(26.2) \qquad\qquad e_t = \sigma_t Z_t,$$

where the random variables Z_t are identically distributed independent variables, usually standard normal variables, and the time series $\{\sigma_t\}$ one models in various ways, most frequently as an AR

[25.2]In other words, we take linear linear approximations to the functions f and g at the places indicated.

or ARMA or ARIMA process; one assumes that Z_t is independent of σ_t.[26.1] Such models are called *autoregressive conditional heteroskedastic*, or ARCH, models, or GARCH (*generalized* ARCH) models.[26.2] Such models were introduced by Robert F. Engle in 1982 in the paper [16]; this paper earned him the Nobel Memorial Prize in Economic Sciences in 2003. Since then, many such models have beed described; a glossary to such models is given by Tim Bellerslev in the paper [3].

To build a GARCH model, on first builds an ARMA or ARIMA model as in formula (26.1), then one estimates the errors e_t in the model; this can be done in the way described in Subsections 9.5 or 10.1. Then one models the variances σ_t of these errors by a linear model:

(26.3)
$$\sigma_t^2 = \omega + \sum_{k=1}^{q'} \alpha_k e_{t-k}^2 + \sum_{l=1}^{p'} \beta_l \sigma_{t-l}^2$$

Note that this is not an ARMA model for the time series of σ_t, since the residuals e_t come from the original time series modeled in equation (26.1) rather than from the sequence of variances σ_t^2. The coefficients ω, α_k, and β_k can be estimated by least square methods or by maximum likelihood methods (the latter give better results according to Engle [16, p. 998]).

26.1 Maximum likelihood for estimate for the coefficients in a GARCH model

We will outline how the maximum likelihood method can be used to determine the model parameters in equation (26.3). Given a time series $\{Y_t\}$. write ψ_t for the information available at time t. This includes all the values of $Y_{t'}$ for $t' \leq t$, and given the model described by formula (26.1), also the values of $e_{t'}$ for $t' \leq t$. At time t before observing Y_t, the information available is ψ_{t-1}. We have

$$\mathrm{E}(Y_t|\psi_{t-1}) = \sum_{k=1}^{p} \phi_k Y_{t-k} - \sum_{k=1}^{q} \theta_k e_{t-k},$$

since

$$\mathrm{E}(e_t|\psi_{t-1}) = \mathrm{E}(\sigma_t Z_t|\psi_{t-1}) = \mathrm{E}(\sigma_t|\psi_{t-1})\,\mathrm{E}(Z_t|\psi_{t-1}) = \mathrm{E}(\sigma_t|\psi_{t-1})\,\mathrm{E}(Z_t) = 0;$$

the second equation holds here since Z_t is independent of σ_t, and the third equation holds since Z_t is also independent of ψ_{t-1}.

In the GARCH model, $e_t = \sigma_t Z_t$ is assumed to be a normal variable with mean 0 and variance σ_t^2; this variance depends on the information ψ_{t-1}. The variance of Y_t conditional on ψ_{t-1} is the same:

$$\mathrm{Var}(Y_t|\psi_{t-1}) = \mathrm{Var}(e_t|\psi_{t-1}) = \sigma_t^2.$$

Thus, the density function of e_t conditional on ψ_{t-1} is

$$f_{e_t|\phi_{t-1}}(x) = \frac{1}{\sqrt{2\pi}\,\sigma_t} \exp\left(-\frac{x^2}{2\sigma_t^2}\right).$$

This is also the conditional likelihood function:

$$L_t(\sigma_t, \hat{e}_t) = \frac{1}{\sqrt{2\pi}\,\sigma_t} \exp\left(-\frac{\hat{e}_t^2}{2\sigma_t^2}\right),$$

[26.1]If one wants to interpret these specifications mathematically, σ_t must also be a random variable. In the model, σ_t will be a function of random variables that assumed numerical values before time t, whereas Z_t is a random variable that assumes a value only at time t.

[26.2]The word *skedastic* or *scedastic* means "related to the variance of statistical errors." Hence *homoskedastic* means having the same finite variance, and *heteroskedastic* means not having the same variance (of errors of a time series).

where we wrote the observed value of $\hat{e}_t$ of the error in place of x. The likelihood function is the product of all conditional likelihood functions for the series of observations of Y_n. The values of $\hat{e}_n$ are not directly observable; they are calculated from the values of Y_n and the model parameters ϕ_k and θ_l in equation (26.1); the values of σ_n are expressed in terms of the model parameters α_k and β_l in equation equation (26.3). The likelihood function is considered as a function of the parameters ω, and α_k and β_l occurring in equation (26.3); these are the parameters to be determined. The model parameters ϕ_k and θ_l in equation (26.1) are assumed to be known at these points. Assuming that observations Y_n were made for times n for $1 \leq n \leq t$, we have

$$L(\omega, \boldsymbol{\alpha}, \boldsymbol{\beta}; \mathbf{Y}) = \prod_{n=1}^{t} \frac{1}{\sqrt{2\pi}\,\sigma_n} \exp\left(-\frac{\hat{e}_n^2}{2\sigma_n^2}\right).$$

In this equation, $\boldsymbol{\alpha} = \langle \alpha_1, \alpha_2, \dots, \alpha_{q'} \rangle$, $\boldsymbol{\beta} = \langle \beta_1, \beta_2, \dots, \beta_{p'} \rangle$, and $\mathbf{Y} = \langle Y_1, Y_2, \dots, Y_t \rangle$, the values for σ_n should be expressed in terms of ω, $\boldsymbol{\alpha}$, and and $\boldsymbol{\beta}$ using equation (26.3), where, at this point, the values of the parameters ω, $\boldsymbol{\alpha}$, and and $\boldsymbol{\beta}$ are yet to be determined. After this, the values of these parameters can be estimating by maximizing the likelihood functions. In doing to, one first takes the logarithm of the likelihood function.

27 The generalized least squares method

For a complex random column vector $\mathbf{Y}$, define the variance of $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^T$ as

$$\mathrm{Var}(\mathbf{Y}) \stackrel{def}{=} \mathrm{E}\left((\mathbf{Y} - \mathrm{E}(\mathbf{Y}))^*(\mathbf{Y} - \mathrm{E}(\mathbf{Y}))\right) = \sum_{k=1}^{n} \mathrm{E}\left(\left|Y_k - \mathrm{E}(Y_k)\right|^2\right).$$

This is a scalar; compare this with the definition of the covariance matrix of $\mathbf{Y}$, which is an $n \times n$ matrix:

$$\mathrm{Cov}(\mathbf{Y}) \stackrel{def}{=} \mathrm{E}\left((\mathbf{Y} - \mathrm{E}(\mathbf{Y}))(\mathbf{Y} - \mathrm{E}(\mathbf{Y}))^*\right)$$

In fact, the variance of $\mathbf{Y}$ is the *trace* of its covariance matrix, where the trace of a square matrix is defined as the sum of its diagonal elements.

27.1 Ordinary least squares

The ordinary least squares method was discussed above, on account of the innovations algorithm in Section 7, especially in the proof of Lemma 7.1. In this subsection we give a description from a different point of view.

Let m and n be a positive integer, A and $m \times n$ matrix with known entries entries and let $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ be a column vector with unkown real entries. We want to determine the unknown entries of $\mathbf{x}$ by measuring the entries of the column vector $A\mathbf{x}$, but these measurements have errors. A mathematical formulation of the problem is the following:

Let $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_m)^T$ and $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)^T$ be a column vectors of random variables. Assume that $\mathrm{E}(\epsilon_k) = 0$ and $\mathrm{E}(\epsilon_k \epsilon_l) = \delta_{kl}\sigma^2$ for all k and l with $1 \leq k, l \leq m$ for some finite $\sigma > 0$. Assume we have the the system of equations

(27.1) $$A\mathbf{x} + \boldsymbol{\epsilon} = \mathbf{Y}.$$

Here we have m equations to determine n unknowns; since the right-hand side of the equations are measured quantities (the measured values of the random variables Y_k), it is likely that these

equations are contradictory. So what we need to do is to find the best estimate for the solutions that the errors $\boldsymbol{\epsilon}$ is these equations is in some sense the least possible. As for the solvability of these equations, the matrix A has to have at least m linearly independent rows, since otherwise we do not have enough equations to find $\mathbf{x}$. In other words, the matrix must have rank n.

The row rank of a matrix is its maximum number of linearly independent rows, and its column rank is equal to the number of linearly independent colums. By a standard theorem of linear algebra, the row rank of a matrix is equal to its column rank, and it is called the rank of the matrix. That is, the rank of A is $\leq \min(m, n)$, since the rank cannot exceed the total number or rows or the total number of columns. So our requerement that the number is linearly independent rows of A be at least n is satisfiable only if $m \geq n$, i.e., if we have at least as many equations as we have unknowns. In a practical situation, we usually have $m > n$, since m is the number of measurements of the random variables Y_i, and we usually want to make more measurements to get a better estimate. The errors $\boldsymbol{\epsilon}$ arise from the errors of the measurements.

In handling equation (27.1), we want to find the best estimate $\hat{\mathbf{x}}$ of the column vector $\mathbf{x}$. A linear estimator for $\mathbf{x}$ is a random column vector $\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n)^T = B\mathbf{Y}$, where B is an $n \times m$ matrix of reals; the entries of the matrix may depend on the entries of the known matrix A, but they must not depend on the components of the unknown vector $\mathbf{x}$ or on the components of the random vector $\mathbf{Y}$. Such an estimator is *unbiased* if $\mathrm{E}(\hat{\mathbf{x}}) = \mathbf{x}$.

We have
$$\mathrm{E}(\hat{\mathbf{x}}) = \mathrm{E}(B\mathbf{Y}) = \mathrm{E}\big(B(\mathbf{Ax} + \boldsymbol{\epsilon})\big) = BA\mathbf{x};$$

the last equation holds since since $\mathrm{E}(\boldsymbol{\epsilon}) = 0$. So, the condition for the estimate $\hat{\mathbf{x}} = BY$ to be unbiased is that

(27.2)
$$BA\mathbf{x} = \mathbf{x}.$$

Writing $\hat{\boldsymbol{\epsilon}} = \mathbf{Y} - A\hat{\mathbf{x}}$, the method of ordinary least squares seeks to minimize the quantity

(27.3)
$$\hat{\boldsymbol{\epsilon}}^T\hat{\boldsymbol{\epsilon}} = (-\hat{\boldsymbol{\epsilon}}^T)(-\hat{\boldsymbol{\epsilon}}) = (A\hat{\mathbf{x}} - \mathbf{Y})^T(A\hat{\mathbf{x}} - \mathbf{Y}).$$

If the matrix $A^T A$ is invertible, then the choice $\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{Y}$ minimizes this expression. First note that this is an unbiased linear estimate with $B = (A^T A)^{-1} A^T$. Indeed $BA\mathbf{x} = (A^T A)^{-1} A^T A\mathbf{x} = \mathbf{x}$, so (27.2) is satisfied.

If A and B are matrices such that the number of columns of A is the same as the number of rows of B, so that the product AB can be formed, then the rows of AB are linear combinations of the rows of B, and its columns are linear combinations of the columns of A. Hence the rank of AB is at most the column rank of A and the row rank of B.

Note that $A^T A$ is an $n \times n$ matrix. For it to be invertible, it has to have rank n. On the other hand, A is an $m \times n$ matrix, and so its rank is $\leq \min(m, n)$. Hence the rank of $A^T A$ is also $\leq \min(m, n)$. Hence, the matrix $A^T A$ can be nonsingular[27.1] only in case $m \geq n$, i.e., if the number of scalar equations given by (27.1) is at least the number of unknown.[27.2]

Next we show that it minimizes $\hat{\boldsymbol{\epsilon}}^T\hat{\boldsymbol{\epsilon}}$:

[27.1]¡*Nonsingular* for a matrix means the same as invertible.
[27.2]The exact condition for the matrix $A^T A$ is that the rank of A be n; this is clear from the discussion above. This means that the system of scalar equations given by (27.1) should contain n independent equations (which equations are then uniquely solvable for $\mathbf{x}$ without errors, i.e., such that $\boldsymbol{\epsilon} = 0$). The solution of these n equations may contradict other equations present in the system, so, usually the whole system of equations are not solvable without errors.

Proof of minimization. Writing $D = (A^T A)^{-1}$, assume that $\hat{\mathbf{x}} = (DA^T + C)\mathbf{Y}$ for some $n \times m$ matrix C. Then, with I being the $m \times m$ identity matrix, we have

$$\begin{aligned}
\hat{\boldsymbol{\epsilon}}^T \hat{\boldsymbol{\epsilon}} &= (-\hat{\boldsymbol{\epsilon}}^T)(-\hat{\boldsymbol{\epsilon}}) = (A\hat{\mathbf{x}} - \mathbf{Y})^T (A\hat{\mathbf{x}} - \mathbf{Y}) \\
&= \mathbf{Y}^T \left(A(DA^T + C) - I \right)^T \left(A(DA^T + C) - I \right) \mathbf{Y} \\
&= \mathbf{Y}^T \left((ADA^T - I) + AC \right)^T \left((ADA^T - I) + AC \right) \mathbf{Y} \\
&= \mathbf{Y}^T (ADA^T - I)^T (ADA^T - I)\mathbf{Y} \\
&\quad + \mathbf{Y}^T (C^T A^T ADA^T - C^T A^T + ADA^T AC - AC + C^T A^T AC)\mathbf{Y};
\end{aligned}$$

in the last equation, we made use of the fact that $D^T = D$; this is because $D = (A^T A)^{-1}$ and so $D^T = \left((A^T A)^T \right)^{-1} = (A^T A)^{-1} = D$. Making use of the fact that $D = (A^T A)^{-1}$, there are cancelations in the second term on the right-hand side, and we obtain that this right-hand side is equal to

$$\begin{aligned}
\mathbf{Y}^T (ADA^T - I)^T &(ADA^T - I)\mathbf{Y} + \mathbf{Y}^T C^T A^T AC\mathbf{Y} \\
&= \mathbf{Y}^T ((ADA^T - I)^T (ADA^T - I)\mathbf{Y} + (AC\mathbf{Y})^T AC\mathbf{Y};
\end{aligned}$$

As $(AC\mathbf{Y})^T AC\mathbf{Y} \geq 0$ if follows that the right-hand side is the minimum when $C = 0$. It is not guaranteed that this is the only minimum, since it is possible that $C\mathbf{Y} = 0$ even if $C \neq 0$ (on the other hand, $AC\mathbf{Y} = 0$ only if $C\mathbf{Y} = 0$, since if $AC\mathbf{Y} = 0$ then $A^T AC\mathbf{Y} = 0$, and the matrix $A^T A$ is nonsingular). $\qquad\square$

The least squares method is attributed to Gauss; he used it to determine the orbit of the dwarf planet Ceres; the method he used is described at the website. The method was first published by Legendre.

Given a vector random variable $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)^T$, we define its variance as

$$\mathrm{Var}(\mathbf{Z}) \overset{def}{=} \sum_{k=1}^{n} \mathrm{Var}(Z_k).$$

It is easy to see that $\mathrm{Var}(\mathbf{Z})$ is the trace of the matrix $\mathrm{Cov}(\mathbf{Z}) \overset{def}{=} \mathrm{E}\left((\mathbf{Z} - \mathrm{E}(\mathbf{Z}))((\mathbf{Z} - \mathrm{E}(\mathbf{Z}))^T \right)$. The linear estimator $\hat{\mathbf{x}}$ is called the *best linear unbiased estimator* if in addition to being unbiased, $\hat{\mathbf{x}} = B\mathbf{Y}$ is such that for any other $n \times m$ matrix B' such that $B'Y$ is an unbiased estimate of $\mathbf{x}$, i.e., such that

(27.4) $$B'A\mathbf{x} = \mathbf{x}$$

in analogy with equation (27.2), we have $\mathrm{Var}(B\mathbf{Y}) \leq \mathrm{Var}(B'\mathbf{Y})$. We have

Theorem 27.1 (Gauss–Markov theorem)**.** *Assume that in equation (27.1) we have $\mathrm{E}(\epsilon_k) = 0$ and $\mathrm{E}(\epsilon_k \epsilon_l) = \delta_{kl} \sigma^2$ for all k and l with $1 \leq k, l \leq m$ for some finite $\sigma > 0$.*[27.3] *If the matrix $A^T A$ is nonsingular, then the best linear unbiased estimator for $\mathbf{x}$ is $\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{Y}$.*

Proof. Let B' be any $n \times m$ matrix. By equation (27.1) we have

$$\mathrm{E}(B'\mathbf{Y}) = \mathrm{E}\left(B'(A\mathbf{x} + \boldsymbol{\epsilon}) \right) = B'A\mathbf{x}.$$

[27.3]These are the same assumptions we made above, just before equation (27.1).

Furthermore,

$$\text{Cov}(B'\mathbf{Y}) = \text{E}\left((B'\boldsymbol{\epsilon})(B'\boldsymbol{\epsilon})^T\right) = \text{E}\left(B'\boldsymbol{\epsilon}(\boldsymbol{\epsilon})^T B'^T\right)$$
$$= B'\,\text{E}\left(\boldsymbol{\epsilon}(\boldsymbol{\epsilon})^T\right)B'^T = B'\sigma^2 I B'^T = \sigma^2 B'B'^T,$$

where I is the $m \times m$ identity matrix. Writing $D = (A^T A)^{-1}$ as before, assume $B' = DA^T + C$ for some $m \times n$ matrix C. In order for the estimate $B'\mathbf{Y}$ to be unbiased, according to equation (27.4) we need to have $CA\mathbf{x} = 0$, since we have $DA^T A\mathbf{x} = (A^T A)^{-1}A^T A\mathbf{x} = \mathbf{x}$. Since the entries C cannot depend on $\mathbf{x}$, this means that we must have $CA\mathbf{x} = 0$ for any $\mathbf{x}$, that is $CA = 0$. Using equation (27.5), and noting that $D^T = D$ as we pointed out before, we have

$$\frac{1}{\sigma^2}\text{Cov}(B'\mathbf{Y}) = B'B'^T = (DA^T + C)(DA^T + C)^T = (DA^T + C)(AD + C^T)$$
$$= DA^T AD + DA^T C^T + CAD + CC^T = (DA^T)(DA^T)^T + CC^T;$$

the last equation holds since $CA = 0$, and so also $A^T C^T = (CA)^T = 0$. Incidentally, $DA^T AD = D$ since $D = (A^T A)^{-1}$, but we did not need to use this. The matrix CC^T is positive semidefinite, and so its trace is nonnegative. Hence, writing $\text{Tr}(G)$ for the trace of a square matrix G, we have

$$\text{Var}(B'\mathbf{Y}) = \text{Tr}\big(\text{Cov}(B'\mathbf{Y})\big) = \sigma^2\,\text{Tr}\left(DA^T(DA^T)^T + CC^T\right)$$
$$= \sigma^2\,\text{Tr}\left(DA^T(DA^T)^T\right) + \text{Tr}\left(CC^T\right) \geq \sigma^2\,\text{Tr}\left(DA^T(DA^T)^T\right) = \text{Var}(DA^T\mathbf{Y}),$$

showing that $DA^T\mathbf{Y}$ is indeed a best linear unbiased estimate for $\mathbf{x}$. $\qquad\square$

The form $(A^T A)^{-1}A^T\mathbf{x}$ of the solution given ty the least squares method is of theoretical interest, and it is not useful for practical calculations. For practical calculations, a factorization of the matrix A into the product of an orthogonal matrix and an upper triangular matrix is used; for details, see [23], in the section on overdetermined systems of linear equations (currently Section 38, pp. 174–184).

27.2   The generalized least squares method

In the generalized least squares method, one wants to find the best estimate $\hat{\mathbf{x}}$ for $\mathbf{x}$ in equation (27.1), where now one drops the assumption that the components of the error vector $\boldsymbol{\epsilon}$ are uncorrelated and have the same variance; instead one assumes that the covariance matrix $\Sigma = \text{Cov}(\boldsymbol{\epsilon})$ is known and is positive definite, i.e., that $\mathbf{c}^T\Sigma\mathbf{c} > 0$ for any $m \times 1$ nonzero column vector $\mathbf{c}$.[27.4] It is still assumed that the $m \times n$ matrix A has rank n. The generalized least squares method was invented by Alexander Aitken.

The generalized least squares problem can be reduced to the ordinary least squares problem. To see how this can be done, first note that the matrix $\text{Cov}(\boldsymbol{\epsilon}) = \Sigma$ being positive definite and symmetric, there is a lower triangular matrix L such that $LL^T = \Sigma$. The factorization LL^T is called the *Cholesky decomposition* or *Cholesky factorization* of the matrix Σ; for a discussion of the Cholesky decomposition, see [23], the section on positive definite matrices (p. 163 in Section 35).

[27.4]We assume that $\boldsymbol{\epsilon}$ is real, so we do not need to use complex conjugation. We have seen above that the covariance matrix is always positive semidefinite, i.e., that $\mathbf{c}^T\Sigma\mathbf{c} \geq 0$ (see Problem 2.1). The assumption that it is positive definite amounts to the same as assuming that it is also nonsingular.

Since Σ is positive definite, it is invertible, and so the matrices L and L^T are also invertible.[27.5]
Multiply equation (27.1) by L^{-1} on the left to obtain

$$(27.6) \qquad\qquad L^{-1}A\mathbf{x} + L^{-1}\boldsymbol{\epsilon} = L^{-1}\mathbf{Y}.$$

Noting that $(L^{-1})^T = (L^T)^{-1}$, we have

$$\mathrm{Cov}(L^{-1}\boldsymbol{\epsilon}) = \mathrm{E}\big(L^{-1}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T(L^T)^{-1}\big) = L^{-1}\,\mathrm{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T)(L^T)^{-1} = L^{-1}\Sigma(L^T)^{-1} = L^{-1}LL^T(L^T)^{-1} = I.$$

Thus, equation (27.6) represents an ordinary least squares problem, showing how to reduce the generalized least squares problem to an ordinary least squares problem.

27.3   Linear regression models and generalized least squares

A linear regression model of a time series $\{Y_t\}$ is an equation of the form

$$Y_t = \alpha_0 + \sum_{k=1}^{m} \alpha_k u_{k,t} + Z_t,$$

where α_i for k with $0 \le k \le m$ are parameters of the model, $u_{k,t}$ are explanatory variables measured at time t, and $\{Z_t\}$ is the residual time series that is not predicted by the model. When fitting a linear regression model to a time series, the residuals will usually be correlated. In this situation, a better model can be obtained by using the generalized least squares method to fit the model parameters instead of using ordinary least squares. This is discussed in [11, §5.4, p. 98] and in [4, §9.4.2, p. 363].

28   Long memory processes

In order to discuss the convergence of a certain series below, we need a convergence criterion not routinely treated in calculus courses.

28.1   The Dirichlet convergence criterion

Theorem 28.1 (Dirichlet convergence criterion). *Let a_k and b_k for $k \ge 1$ be complex numbers such that*

$$(28.1) \qquad\qquad \lim_{k \to \infty} b_k = 0$$

and

$$(28.2) \qquad\qquad \sum_{k=1}^{\infty} |b_k - b_{k+1}| < \infty.$$

Assume that there is a real number B such that

$$(28.3) \qquad\qquad \left| \sum_{k=1}^{N} a_k \right| < B$$

[27.5]Indeed, if for two $n \times n$ matrices A and B the matrix AB is invertible, then both A and B must also be invertible. This is because $\mathrm{rank}(AB) \le \min(\mathrm{rank}\,A, \mathrm{rank}\,B)$, as we pointed out in the second small letter passage in Subsection 27.1.

for all $N \geq 1$. Then the series

$$(28.4) \qquad \sum_{k=1}^{\infty} a_k b_k$$

converges.

This result is the Generalized Dirichlet Convergence Test. In the original version of the Dirichlet Test, instead of (28.2) one assumes that b_k is real and $b_k \geq b_{k+1} > 0$ for all $k \geq 1$. The Alternating Series Test is a consequence of the original version of the Dirichlet Test. Indeed, one obtains the Alternating Series Test if one takes $a_k = (-1)^{k+1}$. We will comment on the role of the Generalized Dirichlet Test in number theory below.

Proof. To show the above result, write

$$c_n = \sum_{k=1}^{n} a_k \qquad (n \geq 0).$$

Then $a_n = c_n - c_{n-1}$, so, given integers M and N with $0 \leq M < N$ we have

$$\sum_{n=M+1}^{N} a_n b_n = \sum_{n=M+1}^{N} (c_n - c_{n-1}) b_n$$

$$= c_N b_{N+1} - c_M b_{M+1} + \sum_{n=M+1}^{N} c_n (b_n - b_{n+1});$$

the last equation can be easily checked by noting that each term in the middle member is matched by exactly one member on the right-hand side. An equation of this type is called partial summation, or Abel rearrangement, named after the Norwegian mathematician Niels Henrik Abel.[28.1]

Therefore

$$(28.5) \qquad \begin{aligned} \left| \sum_{n=M+1}^{N} a_n b_n \right| &\leq |c_N b_{N+1}| + |c_M b_{M+1}| + \sum_{n=M+1}^{N} |c_n| |b_n - b_{n+1}| \\ &\leq B \left(|b_{N+1}| + |b_{M+1}| + \sum_{n=M+1}^{N} |b_n - b_{n+1}| \right) \qquad (0 \leq M < N); \end{aligned}$$

the second inequality follows in view of (28.3). Making $M \to \infty$, the limit of the right-hand side is 0 in view of (28.1) and (28.2). This shows that the series in (28.4) indeed converges. $\qquad \square$

A Dirichlet series is a sum

$$(28.6) \qquad \sum_{n=1}^{\infty} a_n n^{-s},$$

where the coefficients a_n for $n \geq 1$ are given complex numbers. Johann Peter Gustav Lejeune Dirichlet used these eponymous[28.2] series to establish his famous result that if an arithmetic progression with integer terms contains two relatively prime integers then it contains infinitely many prime numbers. Dirichlet considered these series only for real s; somewhat later, Georg Friedrich Bernhard Riemann used them with complex s in his study of prime numbers. The basic convergence result for Dirichlet series is the following:

[28.1]Abel rearrangement was also discussed above in Theorem 15.2, on account of integration by parts for Stieltjes integrals.

[28.2]I.e., series named after him (later, by others), that is, Dirichlet series.

Theorem 28.2 (Region of convergence of Dirichlet series). *If (28.6) converges for $s = s_0$ with some complex s_0, then it also converges for all complex s with $\Re s > \Re s_0$.*

This is a direct convergence of the Generalized Dirichlet Test. Indeed, assume that

$$\sum_{n=1}^{\infty} a_n n^{-s_0}$$

converges. Then

$$\sum_{n=1}^{\infty} a_n n^{-s} = \sum_{n=1}^{\infty} a_n n^{-s_0} \, n^{-(s-s_0)}.$$

Assuming $\Re(s - s_0) > 0$, we have

$$|n^{-(s-s_0)} - (n+1)^{-(s-s_0)}| = \left| \int_n^{n+1} (s - s_0) t^{-(s-s_0)-1} \, dt \right|$$

$$\leq \left| (s - s_0) n^{-(s-s_0)-1} \right| = |s - s_0| n^{-\Re(s-s_0)-1}.$$

Since the series

$$\sum_{n=1}^{\infty} n^{-\Re(s-s_0)-1}$$

is convergent (e.g., by the Integral Test), the Generalized Dirichlet Test implies that the series in (28.6) is also convergent. If we assume that s and s_0 are real, the same conclusion follows also from the original Dirichlet Test.

28.2 The spectrum revisited

Let y_t be observations of the a time for $0 \leq t < N$, and, as in equation (16.1), we describe the series of observations with a trigonometric polynomial:

$$(28.7) \qquad y_t = \sum_{k=0}^{N-1} c_k e^{2ikt\pi/N} \qquad (0 \leq t < N).$$

To simplify the considerations, we will define y_t for all integers t by putting $y_{t+kN} = y_t$ for all $k \in \mathbb{Z}$; then equation (28.7) will be valid for all integers t. As in equation (16.2), we have

$$(28.8) \qquad c_k = \frac{1}{N} \sum_{t=0}^{N-1} y_t e^{-2ikt\pi/N}.$$

This equation is needed only for $0 \leq k < N$, but we will take it to be valid for all integers k, since instead of the range of summation $0 \leq k < N$, we can take any range of N consecutive integers in equation (28.7); cf. equations (14.7) and (14.9).

The spectrum of this time series will be defined analogously to the periodogram defined in equation (16.3) except that we take frequencies in the range $(-\infty, \infty)$, and do not identify frequencies f and $1 - f$:

$$(28.9) \qquad S\left(\frac{k}{N}\right) = N|c_k|^2.$$

28.3 Differencing and the spectrum

If we write $z_t = y_t - y_{t-1}$ for the differenced series, we have

$$z_t = y_t - y_{t-1} = \sum_{k=0}^{N-1} c_k(e^{2ikt\pi/N} - e^{2ik(t-1)\pi/N}) = \sum_{k=0}^{N-1} c_k(1 - e^{-2ik\pi/N})\, e^{2ikt\pi/N};$$

note that the first equation for $k = 0$ makes use of our stipulation above according to which $y_{-1} = y_{N-1}$; without this stipulation, y_{-1} would make no sense. Writing $c_k(z) = c_k(\nabla y)$ for the interpolation coefficients in this equation, and also writing $c_k(y) \stackrel{def}{=} c_k$, this equation shows that

$$c_k(\nabla y) = (1 - e^{-2ik\pi/N})c_k(y).$$

Noting that

$$|1 - e^{-2ik\pi/N}|^2 = (1 - e^{-2ik\pi/N})(1 - e^{2ik\pi/N}) = 1 - e^{-2ik\pi/N} - e^{2ik\pi/N} + 1$$
$$= 2\left(1 - \cos\frac{2k\pi}{N}\right),$$

we obtain for the spectrums with frequency $\nu = k/N$ that that

(28.10) $$S(\nu, \{\nabla y_t\}) = 2(1 - \cos 2\pi\nu)S(\nu, \{y_t\}).$$

Observe that for ν approaching zero, $1 - \cos 2\pi\nu$ has order of magnitude ν^2;[28.3] On the other hand, on the basis of equation (16.7), one does not even expect that $S(\nu) \to 0$ when $\nu \to 0$.[28.4] The expectation is that $S(\nu) \to c$ with some $c \neq 0$. If $S(\nu) \to \infty$, then equation (28.10) suggests that one would need to difference the time series to remedy this situation.

In trying to estimate the amount of differencing needed in a time series, one can try to estimate the order of magnitude of $S(\nu)$ as $\nu \to 0$; call this order ν^{-2d}, meaning that the size of $S(\nu)$ is some bounded multiple of ν^{-2d}. In this case the time series needs to be differenced d times. The case $0 < d < 1$ is especially interesting, and it leads to fractional differencing.[28.5]

28.4 Fractional differencing

The binomial coefficient $\binom{d}{n}$ is defined for any real d and for every integer $n \geq 0$ by the equation

$$\binom{d}{n} \stackrel{def}{=} \prod_{k=0}^{n-1} \frac{d-k}{n-k} \qquad (n \geq 0).$$

Here, for $n = 0$ we have the empty product, which is interpreted as 1. If d is a positive integer and $n \leq d$, then this is the usual binomial coefficient; if d is a positive integer and $n > 0$, then $\binom{d}{n} = 0$,

---

[28.3]Indeed,

$$\lim_{x \to 0} \frac{1 - \cos x}{x^2} = \frac{1}{2}.$$

[28.4]For an observed time series, the frequency $\nu = k/N$ can assume only discrete values, so, strictly speaking, $\nu \to 0$ does not make sense. In a practical sense, however, saying that $S(\nu)$ approaches 0 when ν approaches 0 make sense, since N is expected to be a large integer.

[28.5]Calculating the spectrum of an observed time series is fairly inexpensive with the fast Fourier transform discussed in Section 29.

since then the factor $d - k$ for $n = k$ is 0. For any real d and for any complex z with $|z| < 1$, we have

$$(1 + z)^d = \sum_{n=0}^{\infty} \binom{d}{n} z^n.$$

The radius of convergence of this series is 1. Differencing d times for noninteger d can be interpreted as replacing the observed time series $\{y_t\}$ with

$$(I - B)^d y_t = \sum_{n=0}^{\infty} \binom{d}{n} (-1)^n B^n y_t = \sum_{n=0}^{\infty} \binom{d}{n} (-1)^n y_{t-n};$$

of course, in practice, one cannot take an infinite series here, so one needs to truncate this series at some point, perhaps at $n = 40$.

28.5   Slow decay of autocorrelation

In most stationary processes the autocorrelation decays exponentially; that is one expects that that $|\gamma_n| = O\left(e^{-\alpha n}\right)$ for some positive α.[28.6] A slower decay, such as $\gamma_n \sim cn^{-\lambda}$ for λ with $0 < \lambda < 1$ and $c > 0$ implies that the spectrum is singular (i.e., tends to infinity at frequency 0). In fact, taking $c = 1$ for the sake of simplicity, the spectrum of a stationary time series with autocorrelation coefficients $\gamma(n) = n^{-\lambda}$ for all $n \geq 0$ and λ with $0 < \lambda < 1$ can be written as

$$(28.11) \qquad\qquad S(\nu) = 1 + 2 \sum_{n=1}^{\infty} n^{-\lambda} \cos(2n\nu\pi)$$

according to equation (16.7), First note that the series on the right-hand side is convergent unless ν is an integer. Indeed, given any positive integer K, we have

$$\sum_{n=1}^{K} \cos(2n\nu\pi) = \frac{1}{2}\left(D_K(2\nu\pi) - 1\right) = \frac{1}{2}\left(\frac{\sin(2K+1)\nu\pi}{\sin\nu\pi} - 1\right)$$

according to (13.7), showing that the absolute values of these sums stay under a bound independent of K. Hence the Dirichlet convergence criterion (Theorem 28.1) implies that the series on the right-hand side of equation (28.11) converges unless ν is an integer.

We will see that

$$(28.12) \qquad\qquad \lim_{\nu \searrow 0} \nu^{1-\lambda} S(\nu) = 2 \int_{0}^{\infty} x^{-\lambda} \cos(2\pi x)\,dx.$$

Indeed, let A be a large positive integer, and consider this integral on the interval $[0, A]$. It is important to recall the definition of the Riemann integral for this. The Riemann integral

$$\int_{a}^{b} f(x)\,dx$$

is defined exactly as the Stieltjes integral

$$\int_{a}^{b} f(x)\,dg(x)$$

[28.6]See footnote 21.3 on p. 82 for the definition of the $O(\cdot)$ notation.

for $g(x) = x$. This definition was given in detail in Section 15, so we will not restate the definition here.

Note that the integral

(28.13)
$$\int_0^A x^{-\lambda} \cos(2\pi x)\, dx$$

is a convergent improper integral with a singularity at $x = 0$, so it not Riemann integrable. It is Riemann integrable on the interval $[\epsilon, A]$ for any $\epsilon > 0$. Yet it will be convenient to approximate it with Riemann sums. Noting that the integrand is decreasing on $[0, 1]$, the Riemann sums will still converge to the integral if for ϵ with $0 < \epsilon < 1$ we take the tags ξ_n at the right endpoints of the partition intervals that intersect $[0, \epsilon]$.[28.7] This can be justified as follows.

On the interval $[0, \epsilon]$, the Riemann sum with tags at the right end points of the partition intervals will be less than the integral. The part of the Riemann sum on the interval $[\epsilon, A]$ will approximate the integral on this part, since the Riemann integral exists there. Making $\epsilon \searrow 0$ we can see that the part of the Riemann sum on $[0, \epsilon]$ will tend to zero, and so Riemann sum on the whole interval $[0, A]$ will approximate the integral.

Let ν be a positive real. Writing $N = \lfloor A/\nu \rfloor + 1$, divide the interval $[0, A]$ into intervals N intervals of length ν, except that the last interval may be shorter, so that $x_n = n\nu$ for n with $0 \le n < N$, and $x_N = A$. Pick the tags $\xi_n \in [x_{n-1}, x_n]$ for n with $1 \le n \le N$ such that $\xi_n = x_n$. The norm of the partition

$$P : 0 = x_0 < x_1 < x_2 < \ldots x_N = A$$

is ν. The Riemann sum

(28.14)
$$\sum_{n=1}^{N} \xi_n^{-\lambda} \cos(2\pi \xi_n)\, (x_n - x_{n-1})$$
$$= \sum_{n=1}^{N-1} N(n\nu)^{-\lambda} \cos(2\pi n\nu)\, \nu + A^{-\lambda} \cos(2\pi A)\, (A - x_{N-1})$$
$$= \nu^{1-\lambda} \sum_{n=1}^{N-1} n^{-\lambda} \cos(2\pi n\nu) + A^{-\lambda}\, \cos(2\pi A)(A - x_{N-1})$$

converges to the integral in (28.13) as $\nu \searrow 0$ Since the $0 \le A - x_{N-1} < \nu$, the term after the sum on the right-hand side tends to 0. Making $A \to \infty$, one is tempted to conclude that equation (28.12) follows.

This argument is, however, not correct. Heuristically, one might be tempted to make this conclusion, but a rigorous proof is somewhat delicate. Cauchy might have been forgiven for accepting such an argument as correct,[28.8] We will present a rigorous proof next.

28.6   A rigorous proof of convergence

The conclusion that can be reached by the argument at the end of the last subsection is that

$$\lim_{A\to\infty} \lim_{\nu \searrow 0} \nu^{1-\lambda} \sum_{n=1}^{\lfloor A/\nu \rfloor} n^{-\lambda} \cos(2\pi n\nu) = 2 \int_0^{\infty} x^{-\lambda} \cos(2\pi x)\, dx,$$

[28.7] All but the last of these partition intervals will entirely be included in $[0, \epsilon]$.

[28.8] Cauchy struggled with understanding the limits of continuous functions. While he played a key role in putting analysis on solid foundations, he published three articles "proving" that the limit of continuous functions is continuous – a statement that turns out to be incorrect. See the posting for a discussion.

whereas what we need to show in order to establish (28.12)

$$(28.15) \qquad \lim_{\nu \searrow 0} \lim_{A \to \infty} \nu^{1-\lambda} \sum_{n=1}^{\lfloor A/\nu \rfloor} n^{-\lambda} \cos(2\pi n \nu) = 2 \int_0^\infty x^{-\lambda} \cos(2\pi x)\, dx.$$

In order to establish the second version, some kind of uniform convergence is needed; it is certainly not true that the convergence in (28.11) is uniform in ν. What is in fact true is that the inside limit in (28.15) is uniform. This can be shown by following through the proof of the Dirichlet convergence criterion (Theorem 28.1), so as to obtain a uniform bound in inequality (28.5). The fact that we already know by the Dirichlet test that the series (28.11) converges somewhat simplifies the argument. Using the Dirichlet kernel defined in formula (13.7), we have

$$(28.16) \qquad D_n(2\nu\pi) = 1 + 2 \sum_{k=1}^n \cos 2k\nu\pi = \frac{\sin(2n+1)\nu\pi}{\sin \nu\pi},$$

where the second equation holds if ν is not an integer (so that the denominator is not zero). Hence, for any $M \geq 1$ we obtain

$$\sum_{n=M}^\infty n^{-\lambda} \cos(2n\nu\pi) = \frac{1}{2} \sum_{n=M}^\infty n^{-\lambda} \big(D_n(2\nu\pi) - D_{n-1}(2\nu\pi) \big)$$

$$= -\frac{M^{-\lambda}}{2} D_{M-1}(2\nu\pi) + \frac{1}{2} \sum_{n=M}^\infty \big(n^{-\lambda} - (n+1)^{-\lambda} \big) D_n(2\nu\pi).$$

Noting that $|D_n(2\nu\pi)| \leq |1/\sin \nu\pi|$ for any ν (with ν not an integer), it follows that

$$\left| \sum_{n=M}^\infty n^{-\lambda} \cos(2n\nu\pi) \right| \leq \frac{1}{2 \sin \nu\pi} \left(M^{-\lambda} + \sum_{n=M}^\infty \big(n^{-\lambda} - (n+1)^{-\lambda} \big) \right).$$

$$= \frac{M^{-\lambda}}{\sin \nu\pi} \qquad (0 < \nu < 1)$$

Hence

$$\left| \nu^{1-\lambda} \sum_{n=\lfloor A/\nu \rfloor+1}^\infty n^{-\lambda} \cos(2n\nu\pi) \right| \leq \nu^{1-\lambda} \frac{(\lfloor A/\nu \rfloor + 1)^{-\lambda}}{\sin \nu\pi}$$

$$\leq \nu^{1-\lambda} \frac{(A/\nu)^{-\lambda}}{\sin \nu\pi} \leq A^{-\lambda} \qquad (0 < \nu < 1/2);$$

the third inequality uses the fact that $\sin x/x \geq 2/\pi$ for x with $0 \leq x \leq \pi/2$ (the minimum is reached for $x = \pi/2$, and so $\nu/\sin \nu\pi \leq 1/2$ for ν with $0 \leq \nu < 1/2$. This is sufficient to establish (28.12).

Indeed, to finish the proof of this, let $\epsilon > 0$ be arbitrary, and let A_0 be such that for $A \geq A_0$ and for ν with $0 < \nu < 1/2$ we have

$$\left| \nu^{1-\lambda} \sum_{n=\lfloor A/\nu \rfloor+1}^\infty n^{-\lambda} \cos(2n\nu\pi) \right| \leq \frac{\epsilon}{3}.$$

Let $A_1 \geq A_0$ be such that

$$\left| \int_0^\infty x^{-\lambda} \cos(2\pi x)\, dx - \int_0^{A_1} x^{-\lambda} \cos(2\pi x)\, dx \right| \leq \frac{\epsilon}{3},$$

where A_1 may of course depend on λ; and, given A_1, let $\nu_0 > 0$ be such that for ν with $0 < \nu < \nu_0$, for the Riemann sums in (28.14) (note that $N = \lfloor A/\nu \rfloor + 1$ in these sums) we have

$$\left| \int_0^{A_1} x^{-\lambda} \cos(2\pi x) \, dx - \nu^{1-\lambda} \sum_{n=1}^{\lfloor A_1/\nu \rfloor} n^{-\lambda} \cos(2\pi n\nu) \right| < \frac{\epsilon}{3}.$$

Putting all these together, for ν with $0 < \nu \le \nu_0$ we have

$$\left| \int_0^\infty x^{-\lambda} \cos(2\pi x) \, dx - \nu^{1-\lambda} \sum_{n=1}^\infty n^{-\lambda} \cos(2\pi n\nu) \right| <$$

$$\le \left| \int_0^\infty x^{-\lambda} \cos(2\pi x) \, dx - \int_0^{A_1} x^{-\lambda} \cos(2\pi x) \, dx \right|$$

$$+ \left| \int_0^{A_1} x^{-\lambda} \cos(2\pi x) \, dx - \nu^{1-\lambda} \sum_{n=1}^{\lfloor A_1/\nu \rfloor} n^{-\lambda} \cos(2\pi n\nu) \right|$$

$$+ \left| \nu^{1-\lambda} \sum_{n=1}^{\lfloor A_1/\nu \rfloor} n^{-\lambda} \cos(2n\nu\pi) - \nu^{1-\lambda} \sum_{n=1}^\infty n^{-\lambda} \cos(2n\nu\pi) \right| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon.$$

Since $\epsilon > 0$ was arbitrary, equation (28.12) follows.

28.7   Positivity of the limiting integral

Finally, we will show that the integral on the right-hand side of equation (28.12) is positive. We have

$$\int_0^\infty x^{-\lambda} \cos(2\pi x) \, dx = \sum_{n=0}^\infty \int_n^{n+1} x^{-\lambda} \cos(2\pi x) \, dx.$$

$$= \sum_{n=0}^\infty \int_0^1 (x+n)^{-\lambda} \cos(2\pi x) \, dx.$$

We will show that each of the integrals after the sum is positive; hence the sum is positive. Writing $f(x) = (x+n)^{-\lambda}$ for a fixed n, the function $f(x) - f(x+1/2)$ is decreasing, since its derivative

$$-\lambda \left((x+n)^{-\lambda-1} - (x+n+1/2)^{-\lambda-1} \right)$$

is negative. Hence

$$\left(f(x) - f(1/2 + x) \right) - \left(f(1/2 - x) - f(1-x) \right) > 0 \qquad (0 < x < 1/4).$$

Noting that for any x we have

$$\cos 2\pi x = -\cos 2\pi(1/2 + x) = -\cos 2\pi(1/2 - x) = \cos 2\pi(1-x),$$

we have

$$\int_0^1 (x+n)^{-\lambda} \cos(2\pi x) \, dx = \int_0^1 f(x) \cos(2\pi x) \, dx$$

$$= \int_0^{1/4} \left(\left(f(x) - f(1/2 + x) \right) - \left(f(1/2 - x) - f(1-x) \right) \right) \cos(2\pi x) \, dx > 0,$$

117

since both factors in the integrand on the right-hand side are positive, except for being zero at finitely many points. This shows that the integral in (28.12) is indeed positive.

To conclude, the above considerations show that the time series described at the beginning of Section 28.5 needs to be differenced fractionally at $(1 - \lambda)/2$ times.

28.8 Absolute integrability

When approximating the integral on the right-hand side of equation (28.12) we had to deal with the singularities at 0 and $+\infty$, but we dealt with them in very different ways. When using Riemann sums to approximate the integrals, we could almost totally ignore the singularity at 0 in that the only thing we needed to do is to take the tag at the minimum of the function in the partition interval. On the other hand, we had to exclude the singularity at $+\infty$ by cutting off a neighborhood of infinity from the interval of integration. The reason for this is the very different nature of the singularities. At zero, the absolute value of the integrand is integrable (in fact, the integrand near 0 is positive, so it is its own absolute value), while near infinity, the absolute value of the integrand is not integrable. This makes a big difference in how the integral can be handled; the situation is similar to the difference between absolutely and conditionally convergent series: it is much easier to work with an absolutely convergent series than with a conditionally convergent series. For example, an absolutely convergent series can be rearranged and still have the same sum, while a conditionally convergent series of reals can be made to diverge to $+\infty$ and to $-\infty$, or given any real number c, it can be rearranged so as to converge to r (this is a theorem of Dirichlet).

As we discussed above, the Riemann integral have certain disadvantages, and these disadvantages have been remedied by the Lebesgue integral, discussed above on p. 46 in Subsection 13.4. Lebesgue integrable functions are such that their absolute values are also integrable in the Lebesgue sense. Lebesgue integration extends Riemann integration in a way that simplifies the way one works with integrals, but such a simplification does not seem possible for conditionally convergent integrals such as the one on the right hand side of equation (28.12).

29 The fast Fourier transform

29.1 The discrete Fourier transform

Given a positive integer N, and a sequence $\langle y_n : 0 \leq n < N \rangle$ of complex number, we define its *discrete Fourier transform* as the sequence $\langle \hat{y}_l : 0 \leq l < N \rangle$, where

$$(29.1) \qquad \hat{y}_l = \sum_{n=0}^{N-1} y_n e^{-2iln\pi/N}.$$

According to the discussion in Subsection 14.3, we then have

$$(29.2) \qquad y_n = \frac{1}{N} \sum_{l=0}^{N-1} \hat{y}_l e^{2iln\pi/N};$$

see formulas (14.2) and (14.5) especially. The latter formula is also called the *inverse discrete Fourier transform*. It is often convenient to extend these sequences to all integers integers by putting $y_{n+kN} = y_n$ and $\hat{y}_{n+kN} = \hat{y}_n$ for all $n, k \in \mathbb{Z}$; with this extension, the above formulas remain true for all l and n. These formulas are in complete analogy with the continuous Fourier transform and its inverse described in equations (20.1) and (20.2).

29.2 The fast Fourier transform

The *fast Fourier transform* is a group of algorithms that speeds up the calculation on the discrete Fourier transform by rearranging the order of operations in equation (29.1). As described by this equation, the number of multiplications is about N^2, since $\hat{y}_l$ needs to be computed for all values of l with $0 \leq l < N$; in the rearranged version, the number of multiplications is of the order of magnitude $N \log N$.[29.1] This makes the calculation of the discrete Fourier transform fairly inexpensive in most situations.

In the old days, the time needed to perform a computer algorithms was estimated by the number of multiplications needed to perform the algorithm, since at the time multiplications were fairly time consuming, whereas additions were much faster. Perhaps this is still a reasonable way to estimate the time needed for performing an algorithms, but many things changed in computer technology since then that make this way of estimating computer time fairly inaccurate: pipelining (the different parts of the processor performing several multiplications at the same time, each part of the processor working on different stages of the multiplication; other complex operations can similarly be pipelined), caching (storing frequently used data in a fast and relatively expensive memory before transferring it to the main memory, and parallel processing (several different processors – or cores as they are often called) working on different parts of the problem. Often, the assessment of an algorithm involves suitability for efficient processing using the methods described. For this reason, it is quite a complicated task to write an efficient linear algebra program; they are packages that can adapt to different processors; see e.g. BLAS (Basic Linear Algebra Subprograms).

In its simplest incarnation of the fast Fourier transform, the Cooley–Tukey fast Fourier transform algorithm described by James Cooley and John Tukey in 1965,[29.2] assumes that N is a power of 2 and splits up the calculation in equation (29.1) into two parts according as the subscript n is even or odd:

$$(29.3) \qquad \hat{y}_l = \sum_{n=0}^{N/2-1} y_{2n} e^{-2il(2n)\pi/N} + e^{-2il\pi/N} \sum_{n=0}^{N/2-1} y_{2n+1} e^{-2il(2n)\pi/N}$$
$$= \hat{y}_{l,\text{even}} + e^{-2il\pi/N} \hat{y}_{l,\text{odd}}.$$

The calculation uses recursion, calculating $\hat{y}_l$ by first calculating $\hat{y}_{l,\text{even}}$ and $\hat{y}_{l,\text{odd}}$ in a similar way. Note that these need to be calculated only for l with $0 \leq l < N/2$ since

$$\hat{y}_{l,\text{even}} = \hat{y}_{l+N/2,\text{even}} \qquad \text{and} \qquad \hat{y}_{l,\text{odd}} = \hat{y}_{l+N/2,\text{odd}}.$$

If N is not a power of 2, one can use 0-padding, i.e., extending the sequence $\langle \hat{y}_l : 0 \leq l < N \rangle$ by adding 0s at the end so as to make its length a power of 2. There are variants of the algorithm that work of sequences for any composite N, and other versions that work when N is prime. In most cases, the algorithm runs in time $cN \log N$ for some positive constant c, but if one completely wants to avoid 0-padding, there are some exceptional prime values of N, unlikely to be encountered in practice, for which the algorithm requires N^2 multiplications. In most applications, 0-padding causes no harm.

The inverse discrete Fourier transform described by formula (29.2) can be speeded up the same way by a slight modificationof the fast Fourier transform; only a sign change (from $-$ to $+$ in the exponents) is involved.

[29.1] As always in these notes, log indicates natural logarithm, though it makes no difference in the present context except that the base of logarithm needs to be greater than 1. This is because of the base conversion formula for logarithms: we have $\log_a x = \log_b x / \log_b a$ ($a, b > 0$, $a, b \neq 1$, and $x > 0$).

[29.2] The algorithm was also described by Gauss in an unpublished manuscript dating back to around 1805.

29.3 The number of multiplications needed

Assuming N is a power of 2, we will show that the number of multiplications needed to perform the fast Fourier transform on a sequence of length N so as to calculate all values of $\hat{y}_l$ $(0 \leq l < N)$ is $N \log_2 N$. Using induction, assume this is true for every $M < N$ replacing N, where M is a power of 2. On the right-hand side of (29.3) there is one new multiplication for each value of l $(0 \leq l < N)$ amounting to N multiplications. at most $(N/2) \log_2(N/2)$ multiplications to calculate all values $\hat{y}_{l,\text{even}}$ and at most $(N/2) \log_2(N/2)$ multiplications to calculate all values of $\hat{y}_{l,\text{odd}}$. Thus, the total number of multiplications needed is

$$N + 2(N/2) \log_2(N/2) = N + 2(N/2)(\log_2 N - 1)$$
$$= N + N(\log_2 N - 1) = N \log_2 N.$$

For $N = 1$ no multiplications are needed since in that case the only value of y is y_0, and we have $\hat{y}_0 = y_0$.

30 Representation of band-limited functions

Let f be a square integrable function continuous function such that its Fourier transform $\hat{f}$ is zero outside a the interval $(-\pi, \pi)$.[30.1] In the discussion below, we will omit a rigorous discussion of convergence issues.

In electronic technology, such functions whose Fourier transforms are restricted to a finite interval are called band-limited. They are very important in signal processing, since, as we will see, they can be reproduced exactly by sampling at regular time intervals. Band-limited signals can be produced by analog electronic filters before digital processing (when they can be further filtered). By the Fourier inversion formula (20.2)

$$(30.1) \qquad f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \hat{f}(y) e^{ixy} \, dy.$$

Note that $\hat{f}$ is also square integrable, since the Fourier transform is an isometry according to Subsection 20.2. Represent $\hat{f}$ as a Fourier series on $(-\pi, \pi)$ as

$$(30.2) \qquad \hat{f}(x) = \sum_{n=-\infty}^{\infty} c_n e^{inx},$$

where

$$(30.3) \qquad c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{f}(x) e^{-inx} \, dx = \frac{1}{\sqrt{2\pi}} f(-n) \qquad (-\infty < n < \infty),$$

where the last equation follows from (30.1). Substituting this into equation (30.2), and then into equation (30.1) we obtain

$$(30.4)$$
$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \sum_{n=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} f(-n) \, e^{iny} \, e^{ixy} \, dy$$
$$= \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} f(-n) \int_{-\pi}^{\pi} e^{i(x+n)y} \, dy.$$

[30.1]We could take any other finite interval want to consider only the simplest case.

By making the substitution $t = i(x + k)y$, where t and y are the variables and x is a parameter, we have

$$
\begin{aligned}
(30.5) \qquad \int_{-\pi}^{\pi} e^{i(x+n)y}\, dy &= \frac{1}{i(x+n)} \int_{-\pi i(x+n)}^{\pi i(x+n)} e^t\, dt \\
&= \frac{e^{i\pi(x+n)} - e^{-i\pi(x+n)}}{i(x+n)} = \frac{2\sin\pi(x+n)}{x+n},
\end{aligned}
$$

where the last equation follows from the Euler formula (13.10); for $x = 0$, we take $\sin x / x = 1$ (this makes the right-hand side 2π in case $x + n = 0$; in this case, the integrand on the left-hand side is 1, so this indeed gives the correct result). Substituting this into the above formula, we arrive at

$$
f(x) = \frac{1}{\pi} \sum_{n=-\infty}^{\infty} f(-n) \frac{\sin\pi(x+n)}{x+n}.
$$

Replacing n by $-n$, one may also write

$$
(30.6) \qquad f(x) = \frac{1}{\pi} \sum_{n=-\infty}^{\infty} f(n) \frac{\sin\pi(x-n)}{x-n}.
$$

This formula is called the Whittaker–Shannon interpolation formula[30.2]

30.1 The Nyquist–Shannon sampling theorem

We can interpret formula (30.6) as follows. In the formula

$$
\hat{f}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(y) e^{-ixy}\, dy,
$$

the bandwidth limit $\pm\pi$ corresponds to the value $x = \pm\pi$, then the exponential $e^{i\pi y}$ or $e^{-i\pi y}$ as a function of y has has a period of 2. Thinking of y as time, this corresponds to the frequency $1/2$ per unit time. Then formula formula (30.6) says that if the maximum frequency is $1/2$, then the function f can be perfectly reconstructed by sampling it once at integer times. This explains the Nyquist frequency described in Subsection 16.2 from a mathematical point of view.

30.2 The Poisson summation formula

Assume $|f|$ is integrable on $(-\infty, \infty)$, and let

$$
(30.7) \qquad g(x) = \sum_{k=-\infty}^{\infty} f(x + 2k\pi)
$$

The Fourier series of g is

$$
g(x) = \sum_{n=-\infty}^{\infty} c_n e^{inx},
$$

[30.2]The formula occurs in the works of Whittaker in 1915, in those of Claude Shannon in 1949, but it occurs even earlier in the works of E. Borel in 1898 G. H. Hardy also discovered the formula in 1911.

where

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} g(t) e^{-int}\, dt \qquad (-\infty < n < \infty)$$

according to equations (13.11) and (13.15). For $x = 0$ these give

$$g(0) = \sum_{n=-\infty}^{\infty} c_n = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \int_{-\pi}^{\pi} g(t) e^{-int}\, dt$$

$$= \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \int_{-\pi}^{\pi} \sum_{k=-\infty}^{\infty} f(t + 2k\pi) e^{-int}\, dt$$

$$= \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \int_{-\pi}^{\pi} f(t + 2k\pi) e^{-in(t+2k\pi)}\, dt;$$

in the last step, we interchanged the integration and the sum, and used the equation $e^{-int} = e^{-int}e^{2k\pi} = e^{-in(t+2k\pi)}$ (because $e^{2k\pi i} = 1$). In the last integral we can substitute $x = t + 2k\pi$:

$$g(0) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \int_{(2k-1)\pi}^{(2k+1)\pi} f(x) e^{-inx}\, dx$$

$$= \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} f(x) e^{-inx}\, dx = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} \hat{f}(n).$$

Taking equation (30.7) into account, this gives

(30.8)
$$\sum_{n=-\infty}^{\infty} f(2\pi n) = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} \hat{f}(n).$$

This is called the *Poisson summation formula*. The formula is widely used in number theory, and it has several important modern generalizations.

30.3 Simple properties of the Fourier transform

In order to extend the Poisson summation formula to more general situations, we need the following simple properties of the Fourier transform:

Lemma 30.1. *Let f be a complex-valued integrable function on $\mathbb{R}$, and let α be a real number. If $g(x) = f(x)e^{i\alpha x}$, then $\hat{g}(x) = \hat{f}(x - \alpha)$, and if $h(x) = f(x + \alpha)$ then $\hat{h}(x) = \hat{f}(x)e^{i\alpha x}$. Further, if $k(x) = f(x/\lambda)$ with some $\lambda > 0$, then $\hat{k}(x) = \lambda \hat{f}(\lambda x)$.*

Proof. The proof of these statements consists in simple substitutions in formula (20.1). We have

$$\hat{g}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(y) e^{-ixy}\, dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(y) e^{i\alpha y} e^{-ixy}\, dy$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(y) e^{-i(x-\alpha)y}\, dy = \hat{f}(x - \alpha),$$

and

$$\hat{h}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} h(y) e^{-ixy}\, dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(y + \alpha) e^{-ixy}\, dy$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-ix(t-\alpha)}\, dt = e^{i\alpha x} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-ixt}\, dt = e^{i\alpha x} \hat{f}(x);$$

here, for the third equation, we used the substitution $t = y + \alpha$. Finally

$$\hat{k}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(y/\lambda)e^{-ixy}\,dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t)e^{-ix\lambda t}\lambda\,dt = \lambda\hat{f}(\lambda x),$$

where the second equation was obtained by making the substitution $t = y/\lambda$.

$\square$

Using this Lemma with t replacing α, we can restate the Poisson summation formula (30.8) as

$$\sum_{n=-\infty}^{\infty} f(2\pi n)e^{2i\pi nt} = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} \hat{f}(n-t),$$

and as

$$\sum_{n=-\infty}^{\infty} f(2\pi n + t) = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} \hat{f}(n)e^{int}$$

for any real t. Using the part of the lemma for the Fourier transform of $k(x)$, we can also make a scale change in these formulas; for example, with $\lambda = 2\pi$, the former of these formulas can be rewritten as

$$\sum_{n=-\infty}^{\infty} f(n)e^{2\pi int} = \sqrt{2\pi} \sum_{n=-\infty}^{\infty} \hat{f}(2\pi(n-t)).$$

Restating this with $x = -2\pi t$, we obtain

$$(30.9) \qquad \sum_{n=-\infty}^{\infty} f(n)e^{-inx} = \sqrt{2\pi} \sum_{n=-\infty}^{\infty} \hat{f}(2\pi n + x).$$

30.4   Aliasing

Equation (30.9) can also be written as

$$(30.10) \qquad \sum_{n=-\infty}^{\infty} \hat{f}(x + 2\pi n) = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} f(n)e^{-inx} = \sum_{n=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} f(-n)e^{inx}$$

If we assume that $\hat{f}(x) = 0$ for $|x| \geq \pi$, then this equation becomes identical to what is expressed by equations (30.2) and (30.3) together.[30.3] Recall that in those equations this assumption was indeed make. Our aim here is to study how f can be reconstructed from its sampled values at integer arguments, i.e., from the values $f(n)$ for $n \in \mathbb{Z}$.

From this point on, we can mimic the derivation of formula (30.6), but the result we obtain will be different. Write

$$(30.11) \qquad G(x) = \sum_{n=-\infty}^{\infty} \hat{f}(x + 2\pi n)$$

and write

$$(30.12) \qquad F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} G(y)e^{ixy}\,dy.$$

[30.3]In the former of these two equations, we assumed that x is in the interval $(-\pi, \pi)$. It is unnecessary to make this assumption here, since both sides of equation (30.10) are periodic with period 2π.

Note that if $\hat{f}(x) = 0$ for $|x| \geq \pi$ then $F(x) = f(x)$ according to equation (30.1). Observing that

$$G(y) = \sum_{n=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} f(-n) e^{iny}$$

in view of equation (30.10), and poceeding similarly as in equation (30.4), we obtain

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} G(y) e^{ixy}\, dy. = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \sum_{n=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} f(-n)\, e^{iny}\, e^{ixy}\, dy$$

$$= \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} f(-n) \int_{-\pi}^{\pi} e^{i(x+n)y}\, dy = \frac{1}{\pi} \sum_{n=-\infty}^{\infty} f(-n) \frac{\sin \pi(x+n)}{x+n},$$

where the last equation used formula (30.5). Replacing n by $-n$ in the summation on the right-hand side, we obtain an equation similar to (30.6):

$$F(x) = \frac{1}{\pi} \sum_{n=-\infty}^{\infty} f(n) \frac{\sin \pi(x-n)}{x-n}.$$

The problem is that $F(x)$ is usually different from $f(x)$.

Indeed, according to formulas (30.11) and (30.12) we have

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \sum_{n=-\infty}^{\infty} \hat{f}(y + 2\pi n) e^{ixy}\, dy = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} \int_{-\pi}^{\pi} \hat{f}(y + 2\pi n) e^{ixy}\, dy.$$

According to the Fourier inversion formula (20.2) we have

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{f}(t) e^{ixt}\, dt = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} \int_{-\pi+2n\pi}^{\pi+2n\pi} \hat{f}(t) e^{ixt}\, dt$$

$$= \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} \int_{-\pi}^{\pi} \hat{f}(y + 2\pi n) e^{ix(y+2\pi n)}\, dy$$

$$= \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} e^{2i\pi xn} \int_{-\pi}^{\pi} \hat{f}(y + 2\pi n) e^{ix(y+2\pi n)}\, dy;$$

to obtain the third equation, we made the substitution $y = t - 2n\pi$. Hence

$$F(x) - f(x) = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} (1 - e^{2i\pi xn}) \int_{-\pi}^{\pi} \hat{f}(y + 2\pi n) e^{ixy}\, dy.$$

Here the term for $n = 0$ is 0, but the other terms are not, and they represent the distortions added to the original signal

30.5 Anti-aliasing filter

Aliasing is a real issue for engineering both in image and in audio processing. Anti-aliasing filters are used to eliminate frequencies exceeding the Nyquist frequency. In audio processing, an analog filter may be applied to the incoming audio signal before analog-to-digital conversion; another filter may be used to prevent the distortions in the out-of-band frequencies to enter the analog signal. Digital cameras also use anti-aliasing filters. These can use various techniques, such as *birefringent*[30.4]

[30.4]As in by-refringent, i.e., doubly refracting materials. These are materials that have refractive index depending on the polarization and the direction of the incoming light.

materials that spread out the image of a single point to several (usually four) nearby points, thereby cutting down on high spacial frequencies. Other techniques involve vibrating the optical sensor so as to blur features of the image exceeding the Nyquist frequency (image processing is similar to audio processing, but the image is represented in two spacial directions).

31 Solutions to problems

Solution of Problem 2.1. Let $\mathbf{X} = (X_1, X_2, \ldots X_n)^T$ a random column vector; without loss of generality, we may assume that $\mathrm{E}(X_k) = 0$ for each k with $1 \leq k \leq n$. Writing $A = (a_{ij})$ for its covariance matrix, we have

$$a_{ij} = \mathrm{E}(X_i X_j).$$

Now let $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ be an arbitrary n-dimensional column vector. We have

$$\mathbf{x}^T A \mathbf{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i a_{ij} x_j = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i \, \mathrm{E}(X_i X_j) \, x_j = \mathrm{E}\Big(\sum_{i=1}^{n} \sum_{j=1}^{n} x_i X_i \, x_j X_j \Big)$$

$$= \mathrm{E}\left(\Big(\sum_{i=1}^{n} x_i X_i \Big)^2 \right) \geq 0.$$

This shows that A is indeed positive semidefinite.

Note. One can formulate this argument also in matrix form. Assuming, as before, that $\mathrm{E}(\mathbf{X}) = 0$, the covariance matrix of $\mathbf{X}$ is $A = \mathrm{E}(\mathbf{X}\mathbf{X}^T)$. Hence, given an arbitrary n-dimensional column vector $\mathbf{x}$, we have

$$\mathbf{x}^T A \mathbf{x} = \mathbf{x}^T \, \mathrm{E}(\mathbf{X}\mathbf{X}^T) \mathbf{x} = \mathrm{E}(\mathbf{x}^T \mathbf{X}\mathbf{X}^T \mathbf{x}) = \mathrm{E}\big((\mathbf{x}^T \mathbf{X})(\mathbf{X}^T \mathbf{x}) \big) = \mathrm{E}\big((\mathbf{X}^T \mathbf{x})^T (\mathbf{X}^T \mathbf{x}) \big).$$

Note that $\mathbf{X}^T \mathbf{x}$ is the product of a $1 \times n$ matrix and an $n \times 1$ matrix, so it is a 1×1 matrix, i.e., it is a scalar. Hence it is its own transpose; that is $(\mathbf{X}^T \mathbf{x})^T = \mathbf{X}^T \mathbf{x}$. Thus,

$$\mathbf{x}^T A \mathbf{x} = \mathrm{E}\big((\mathbf{X}^T \mathbf{x})^T (\mathbf{X}^T \mathbf{x}) \big) = \mathrm{E}\big((\mathbf{X}^T \mathbf{x})(\mathbf{X}^T \mathbf{x}) \big) = \mathrm{E}\left((\mathbf{X}^T \mathbf{x})^2 \right) \geq 0,$$

as we wanted to show.

Solution of Problem 4.1. The characteristic equation of the recurrence equation $y_t = y_{t-1} + y_{t-2}$ is $1 = \zeta + \zeta^2$, i.e., is $\zeta^2 + \zeta - 1 = 0$. The solutions of this equation are

$$\zeta_1 = \frac{-1 + \sqrt{5}}{2} = \left(\frac{1 + \sqrt{5}}{2} \right)^{-1} \quad \text{and} \quad \zeta_2 = \frac{-1 - \sqrt{5}}{2} = \left(\frac{1 - \sqrt{5}}{2} \right)^{-1};$$

the easiest way to see these equations is by noting that $\zeta_1 \zeta_2 = -1$. Thus, the general solution of the above recurrence equation is

$$y_t = C_1 \left(\frac{1 + \sqrt{5}}{2} \right)^t + C_2 \left(\frac{1 - \sqrt{5}}{2} \right)^t.$$

The initial conditions $y_0 = 0$ and $y_1 = 1$ lead to the equations

$$C_1 + C_2 = 0$$

and

$$C_1 \frac{1+\sqrt{5}}{2} + C_2 \frac{1-\sqrt{5}}{2} = 1.$$

It is easy to solve these equations. Multiplying the first equation by $1/2$ and subtracting it from the second equation, we obtain

$$\frac{\sqrt{5}}{2}(C_1 - C_2) = 1,$$

that is

$$C_1 - C_2 = \frac{2}{\sqrt{5}},$$

Adding the first equation to this, we obtain $2C_1 = 2/\sqrt{5}$, or else $C_1 = 1/\sqrt{5}$. Substituting this into the first equation, we obtain $C_2 = -1/\sqrt{5}$. With these values for C_1 and C_2, the formula for y_t gives

$$y_t = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^t - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^t.$$

Solution of Problem 4.2. The difference operator

$$(B-3)^3$$

will lower the degree of the polynomial in the first term to 0 (i.e., it will change the term into $c \cdot 3^{-t}$ with a nonzero c), while it will not change the degrees of the other polynomials. The difference operator

$$(B-2)^5$$

will annihilate the second term, while it will not change the degrees of the polynomials in the other terms. Finally, the difference operator

$$(B-5)^3$$

will annihilate the third term, while it will not change the degrees of the polynomials Hence the product of these differential operators,

$$(B-3)^3(B-2)^5(B-5)^3$$

will change the first term into $c \cdot 3^{-t}$ with a nonzero c, while it will annihilate the second and the third terms.

This argument can be used to show that if

$$c_1 t^3 \cdot 3^{-t} + c_2 t^4 \cdot 2^{-t} + c_3 t^2 \cdot 5^{-t} \equiv 0,$$

then we must have $c_1 = 0$. Similar arguments can be used to show that we must also have $c_2 = 0$ and $c_3 = 0$; hence the terms $t^3 \cdot 3^{-t}$, $t^4 \cdot 2^{-t}$, and $t^2 \cdot 5^{-t}$ are linearly independent.

Solution of Problem 5.1. Let A be the event $A = (X \neq 0)$; further, let $A_1 = (X \geq 1)$, and for $n > 1$ let $A_n = \big(1/(n-1) > X \geq 1/n\big)$. We have

$$A = \bigcup_{n=1}^{\infty} A_n.$$

126

As the events $A_m \cap A_n = 0$ if $m \neq n$, this implies that

$$P(A) = \sum_{n=1}^{\infty} P(A_n).$$

As $P(A) > 0$, there is an $n \geq 1$ such that $P(A_n) > 0$. With this n we have

$$E(X) \geq P(A_n) \cdot \frac{1}{n} > 0.$$

Solution of Problem 5.2. We may assume that $P(X \neq 0) > 0$, since otherwise $E(XY) = 0$, so the inequality to be proved clearly holds. Then, according to Problem 5.1, $E(X^2) > 0$. Let λ be an arbitrary real number. Then, again by Problem 5.1, the equation

$$E\big((\lambda X + Y)^2\big) = 0$$

can hold only if $\lambda X + Y = 0$ almost surely. As $X \neq 0$ with positive probability, this equation can only hold for a single value of λ.

Now,

$$E\big((\lambda X + Y)^2\big) = \lambda^2 \, E(X^2) + 2\lambda \, E(XY) + E(Y^2).$$

Considering

$$\lambda^2 \, E(X^2) + 2\lambda \, E(XY) + E(Y^2) = 0$$

as a quadratic equation for λ with the various expectation as coefficients,[31.1] this equation has at most one real solution, Hence its discriminant cannot be positive. That is,

$$\big(2 \, E(XY)\big)^2 - 4 \, E(X^2) \, E(Y^2) \leq 0.$$

Rearranging this, we obtain the inequality to be proved.

Solution of Problem 5.3. We have

$$\mathrm{Corr}(X, Y) = \frac{E\Big((X - E(X))(Y - E(Y))\Big)}{\sqrt{E\Big((X - E(X))^2\Big) E\Big((Y - E(Y))^2\Big)}}.$$

This is between -1 and 1 in view of Schwarz's inquality (cf. Problem 5.2).

Solution of Problem 7.1. These equations, properly arranged, give us a way to evaluate the coefficients $\psi_{n,t}$ and the moments $E(e_t^2)$. Let $t \geq 0$ and n with $0 \leq n \leq t$ be integers, and assume $\psi_{n',t'}$ have been calculated for all pairs (n', t') such that $0 \leq t' < t$ and $0 \leq n' \leq t'$ or $t' = t$ and $n < n' \leq t$; also assume that $E(e_k^2)$ has been calculated for all k with $0 \leq k < t$.

We can start out this calculation in case $t = 0$ by noting that

$$E(e_0^2) = E(Y_0^2)$$

according to equation (7.10) with $t = 0$. If $t > 0$ then $\psi_{t,t}$ can be calculated from equation (7.11) with $n = t$, since the only term on the right-hand side involves $l = 0$, and $\phi_{n-l,n} = \phi_{0,0} = 1$ in this case according to equation (7.7). That is,

$$\psi_{t,t} = E(Y_t Y_0) / E(e_0^2).$$

[31.1]This equation is a genuine quadratic equation, since $E(X^2) \neq 0$, that is, the coefficient of λ^2 is not zero, according to what we said above.

If $0 < n < t$ then $\psi_{n,t}$ can be calculated from the same equation (7.11), since for all the coefficients all the quantities on the right-hand side are known except for the $\psi_{n,t}$, which occurs for $l = t - n$ as part of the term

$$\psi_{0,t-n}\psi_{n,t}\,\mathrm{E}(e_{t-n}^2) = \psi_{n,t}\,\mathrm{E}(e_{t-n}^2);$$

the equation here holds in view of equation (7.7).[31.2] That is,

$$\psi_{n,t} = \big(\mathrm{E}(e_{t-n}^2)\big)^{-1}\Big(\mathrm{E}(Y_t Y_{t-n}) - \sum_{l=0}^{t-n-1}\psi_{t-n-l,t-n}\psi_{t-l,t}\,\mathrm{E}(e_l^2).\Big)$$

Finally, for $n = 0$ we have $\psi_{n,t} = 1$ according to equation (7.7).

As the final step, we can use equation (7.10) to evaluate $\mathrm{E}(e_t^2)$:

$$\mathrm{E}(e_t^2) = \mathrm{E}(Y_t^2) - \sum_{l=0}^{t-1}\psi_{t-l,t}^2\,\mathrm{E}(e_l^2).$$

Solution of Problem 13.1. We have

$$D_n(t)\sin\frac{1}{2}t = \sin\frac{1}{2}t + \sum_{k=1}^{n} 2\cos kt\,\sin\frac{1}{2}t$$

$$= \sin\frac{1}{2}t + \sum_{k=1}^{n}\Big(\sin\Big(k+\frac{1}{2}\Big)t - \sin\Big(k-\frac{1}{2}\Big)t\Big) = \sin\Big(n+\frac{1}{2}\Big)t\,;$$

the second equation uses the fourth equation in (13.3), and the third equation results by cancelations.[31.3]

Solution of Problem 13.2. According to equations (13.4) we have

$$(31.1) \qquad a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} x\cos nx\,dx = 0$$

for $n \geq 0$. The equation here holds since the integrand is an odd function[31.4] so the integral on $[-\pi, 0]$ cancels the integral on $[0, \pi]$. Further, by integration by parts we obtain

$$(31.2) \qquad \begin{aligned} b_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} x\sin nx\,dx = -\frac{1}{\pi}x\,\frac{\cos nx}{n}\Big|_{x=-\pi}^{x=\pi} + \frac{1}{\pi}\int_{-\pi}^{\pi}\frac{\cos nx}{n}\,dx \\ &= -\frac{1}{\pi}x\,\frac{\cos nx}{n}\Big|_{x=-\pi}^{x=\pi} + \frac{1}{\pi}\frac{\sin nx}{n^2}\Big|_{x=-\pi}^{x=\pi} = -\frac{2(-1)^n}{n}\,; \end{aligned}$$

[31.2] For calculating $\psi_{n,t}$ we need to assume that $\mathrm{E}(e_{t-n}^2) \neq 0$. However, in the case of $\mathrm{E}(e_{t-n}) = 0$ we do not need to do any calculations, since in this case $e_{t-n} = 0$ almost surely according to Problem 5.1, and so $\psi_{n,t}$ occurs with coefficient 0 or almost surely 0 in the above equations, and so we can take $\psi_{n,t}$ to be anything (the best is to take $\psi_{n,t} = 0$ in this case.

[31.3] That is, the sum telescopes, or collapses. A telescoping or collapsing sum is a sum of the type

$$\sum_{k=1}^{n}(a_{k+1} - a_k) = (a_2 - a_1) + (a_3 - a_2) + (a_4 - a_3) + \ldots + (a_{n+1} - a_n) = a_{n+1} - a_1.$$

[31.4] The function f is odd if $f(-x) = -f(x)$.

128

the last equation holds because $\cos n\pi = (-1)^n$ and $\sin n\pi = 0$ for integer n. Hence the Fourier series of $f(x)$ is

$$-\sum_{n=1}^{\infty} \frac{2(-1)^n}{n} \sin nx.$$

Solution of Problem 13.3. With $f(x)$ as in Problem 13.2, we have

$$\frac{1}{\pi} \int_{-\pi}^{\pi} \big(f(x)\big)^2 \, dx = \frac{1}{\pi} \int_{-\pi}^{\pi} x^2 \, dx = \frac{1}{\pi} \frac{2\pi^3}{3} = \frac{2\pi^2}{3}.$$

Further, in Problem 13.2 a_n and b_n are given by equations (31.1) and (31.2); using these equations, we have

$$|a_0|^2 + \sum_{n=1}^{\infty} \big(|a_n|^2 + |b_n|^2\big) = \sum_{n=1}^{\infty} \frac{4}{n^2}.$$

According to equation (13.17), the right-hand sides of the last two displayed equations are equal, establishing equation (13.18).

Solution of Problem 14.1. Assume the polynomials $P_1(z)$ and $P_2(z)$ are different. Then

$$P(z) \stackrel{def}{=} P_2(z) - P_1(z)$$

is a nonzero polynomial of degree less than N such that $P(z_k) = 0$ for all k with $1 \le k \le N$. Since the numbers z_k are distinct, this is a contradiction, since a nonzero polynomial of degree less than N cannot have N zeros.

Solution of Problem 15.1. Let

$$P : -1 = x_0 < x_1 < x_2 < \ldots < x_n = 1$$

be a partition and let $\xi_i \in [x_{i-1}, x_i]$ be a tag for each i with $1 \le i \le n$. Let $k = k(P)$ with $1 \le k \le n$ be such that $x_i < 0$ for $i < k$ and $x_i \ge 0$ for $i \ge k$; clearly, k depends on the partition P. Then

$$g(x_i) - g(x_{i-1}) = \begin{cases} 1 & \text{if } i = k, \\ 0 & \text{if } i \ne k. \end{cases} \qquad (1 \le i \le n).$$

Hence

$$S(P) = \sum_{i=1}^{n} f(\xi_i)\big(g(x_i) - g(x_{i-1})\big) = f(\xi_k) = f(\xi_{k(P)});$$

$S(P)$ depends on also on the tags, not just on P, but this dependence is not indicated. Making $\|P\| \to 0$, we have $\xi_{k(P)} \to 0$. Since f is continuous at 0, we have

$$\int_{-1}^{1} f(x) \, dg(x) = \lim_{\|P\| \to 0} S(P) = \lim_{\|P\| \to 0} f(\xi_{k(P)}) = f(0).$$

Solution of Problem 15.2. Formula (15.3) can be written with a sum instead of a Stieltjes integral as

(31.3)
$$\tilde{f}(x) = \frac{1}{N} \sum_{k=0}^{N-1} f(x_k) D_M(x - x_k).$$

129

Solution of Problem 15.3. Instead of substituting the coefficients a_n and b_n from equations (13.4) into equation (13.5), we now substitute equations (14.13) into (14.14), the calculations given in formula (13.8) can be repeated with only minor changes:

(31.4)
$$f(x_n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) \, d\omega_N(y) + \frac{1}{\pi} \sum_{k=1}^{M} \int_{-\pi}^{\pi} f(y)(\cos ky \, \cos kx_n + \sin ky \, \sin kx_n) \, d\omega_N(y)$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) \left(1 + 2 \sum_{k=1}^{M} \cos k(y - x_n) \right) d\omega_N(y) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) D_M(y - x_n) \, d\omega_N(y).$$

It is probably best to stop at this point, and not pursue the rest of the calculations in formula (13.8) since the next step is a change of variable in the integral, and to do this, we would need to use Theorem 15.3, and the expression we obtain that way would result in some complications. Since $D(x) = D(-x)$ for all x, the above equation can also be written as

$$f(x_n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) D_M(x_n - y) \, d\omega_N(y).$$

This establishes equation (15.4).

Solution of Problem 15.4. Let $N = 2M$, and define b_M in analogy with the second equation in (14.13) as

$$b_M \overset{\text{def}}{=} \frac{2}{N} \sum_{n=0}^{N-1} f(x_n) \sin M(x_n - x_0)$$

According to equation (14.1), we have $M(x_n - x_0) = n\pi$, and so equation (14.9) implies

$$b_M = \frac{2}{N} \sum_{n=0}^{N-1} f(x_n) \sin n\pi = 0$$

since $\sin n\pi = 0$ for all integers n, we can rewrite formula (14.18) as

$$f(x_n) = \frac{a_0}{2} + \sum_{k=1}^{M-1} (a_k \cos kx_n + b_k \sin kx_n)$$

$$+ \frac{1}{2} \Big(a_M \cos\big(M(x_n - x_0)\big) + b_M \sin\big(M(x_n - x_0)\big) \Big) \qquad (0 \le n < N).$$

Then, similarly as in equation (31.4) we have

$$f(x_n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) \, d\omega_N(y) + \frac{1}{\pi} \sum_{k=1}^{M-1} \int_{-\pi}^{\pi} f(y)(\cos ky \, \cos kx_n + \sin ky \, \sin kx_n) \, d\omega_N(y)$$

$$+ \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) \Big(\cos\big(M(y - x_0)\big) \, \cos M(x_n - x_0)$$

$$+ \sin M(y - x_0) \, \sin\big(M(x_n - x_0)\big) \Big) \, d\omega_N(y)$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) \left(1 + 2 \sum_{k=1}^{M-1} \cos k(y - x_n) + \cos M(y - x_n) \right) d\omega_N(y)$$

130

This formula can be written as

$$f(x_n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) D_M^{\mathrm{mod}}(y - x_n)\, d\omega_N(y),$$

where

$$D_M^{\mathrm{mod}}(t) \overset{def}{=} 1 + 2 \sum_{k=1}^{M-1} \cos kt + \cos Mt = D_{M-1}(t) + \cos Mt$$

is the modified Dirichlet kernel.[31.5]

Solution of Problem 17.1. We will only consider the case when V is an inner product space over $\mathbb{C}$, since the proof for that case also works when V is an inner product space over $\mathbb{R}$, except that in this latter case complex conjugation has no effect. The proof is similar to the one given in the solution of Problem 5.2, except that taking complex inner products causes minor additional complications.

We may assume that $\langle x, y \rangle \neq 0$, since otherwise the inequality to be proved clearly holds; then we also have $x \neq 0$. Let λ be a complex number. Then, by Clause (a) of Definition 17.1 of inner product, we have

$$\langle \lambda x + y, \lambda x + y \rangle \geq 0,$$

and equation here holds only if $\lambda x + y = 0$. Since we assumed that $x \neq 0$, this equation can only hold for a single value of λ if at all. Hence

(31.5)
$$\begin{aligned}
0 \leq \langle \lambda x + y, \lambda x + y \rangle &= \langle \lambda x, \lambda x \rangle + \langle \lambda x, y \rangle + \langle y, \lambda x \rangle + \langle y, y \rangle \\
&= \lambda^* \lambda \langle x, x \rangle + \lambda^* \langle x, y \rangle + \lambda \langle y, x \rangle + \langle y, y \rangle \\
&= |\lambda|^2 \langle x, x \rangle + 2\Re\big(\lambda \langle x, y \rangle^*\big) + \langle y, y \rangle;
\end{aligned}$$

the third equation holds since $\lambda^* \lambda = |\lambda|^2$, and, with $z = \lambda \langle y, x \rangle = \lambda \langle x, y \rangle^*$, we have $z^* = \lambda^* \langle x, y \rangle$ according to Clause (b) of Definition 17.1, and $z^* + z = 2\Re z$, where the $\Re z$ denotes the real part of z. Let

$$\lambda_0 = \frac{|\langle x, y \rangle|}{\langle x, y \rangle^*},$$

and put $\lambda = \rho \lambda_0$, where ρ is an arbitrary real (recall that we assumed that $\langle x, y \rangle \neq 0$). Then $|\lambda_0| = 1$ and so $|\lambda|^2 = \rho^2$. Further, the expression

$$\lambda \langle x, y \rangle^* = \rho \lambda_0 \langle x, y \rangle^* = \rho |\langle x, y \rangle|$$

is real, and so $\Re\big(\lambda \langle x, y \rangle^*\big) = \rho |\langle x, y \rangle|$. Thus, inequality (31.5) becomes

(31.6)
$$\rho^2 \langle x, x \rangle + 2\rho |\langle x, y \rangle| + \langle y, y \rangle \geq 0.$$

According to what we said about the former inequality, we have equality here for at most one real value of ρ.[31.6] Hence the equation

$$\rho^2 \langle x, x \rangle + 2\rho |\langle x, y \rangle| + \langle y, y \rangle = 0.$$

[31.5]The modified Dirichlet kernel is often denoted as $D_M^*(t)$; we avoided this notation, since we use the asterisk to denote complex conjugate.

[31.6]Saying that ρ is real is important here, since this inequality does not even have to hold if ρ is not real. This inequality is a consequence of inequality (31.5) only for real ρ. This is because the equation $\Re \lambda \langle x, y \rangle^* = \rho |\langle x, y \rangle|$ holds only for real ρ.

is a quadratic equation for ρ with real coefficients (recall that $\langle x, x \rangle \neq 0$ by Clause (a) of Definition 17.1 of inner product, since $x \neq 0$). that has at most one real solution. Hence its discriminant cannot be positive. That is,

$$\big(2\langle x, y \rangle\big)^2 - 4\langle x, x \rangle \langle y, y \rangle \leq 0.$$

Rearranging this, we obtain the inequality to be proved.

This solution can be greatly shortened by taking

$$\lambda = -\frac{|\langle x, y \rangle|^2}{\langle x, x \rangle \langle x, y \rangle^*} = -\frac{\langle x, y \rangle \langle x, y \rangle^*}{\langle x, x \rangle \langle x, y \rangle^*} = -\frac{\langle x, y \rangle}{\langle x, x \rangle}$$

in inequality (31.5). Indeed, this choice corresponds to the choice

$$\rho = -\frac{|\langle x, y \rangle|}{\langle x, x \rangle},$$

which is the value of ρ for which the left-hand side of inequality (31.6) assumes its minimum. Such a shortening is, however, no real simplification, since it is achieved by skipping the explanation why this choice of λ is taken.

Solution of Problem 17.2. As in the solution of Problem 17.1, we assume that V is an inner product space over $\mathbb{C}$. We have

$$\begin{aligned}
(\|x\| + \|y\|)^2 &= \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 \geq \langle x, x \rangle + 2|\langle x, y \rangle| + \langle y, y \rangle \\
&\geq \langle x, x \rangle + 2\Re(\langle x, y \rangle) + \langle y, y \rangle = \langle x, x \rangle + \langle x, y \rangle + \langle x, y \rangle^* + \langle y, y \rangle \\
&= \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle = \langle x + y, x + y \rangle = \|x + y\|^2;
\end{aligned}$$

here the first inequality follows from Schwarz's inequality, established in the solution of Problem 17.1.

Solution of Problem 17.3. We have

$$\lim_{n \to \infty} \big|\langle g, f_n - f \rangle\big| \leq \lim_{n \to \infty} \big(\|g\| \, \|f_n - f\|\big)^{1/2}$$

according to Schwarz's inequality. The limit on the right is 0 in view of our assumptions. This establishes the assertion to be proved.

Solution of Problem 17.4. We need to show that $\langle \cdot, \cdot \rangle$ satisfies Clause (b) in Definition 17.1, and that it satisfies Clause (c) also for complex α in the same definition. We can see the former as follows:

$$\begin{aligned}
\langle g, f \rangle &= \langle g, f \rangle_{\mathbb{R}} + i\langle ig, f \rangle_{\mathbb{R}} = \langle f, g \rangle_{\mathbb{R}} + i\langle f, ig \rangle_{\mathbb{R}} \\
&= \langle f, g \rangle_{\mathbb{R}} + i\langle if, i^2 g \rangle_{\mathbb{R}} = \langle f, g \rangle_{\mathbb{R}} + i\langle if, -g \rangle_{\mathbb{R}} \\
&= \langle f, g \rangle_{\mathbb{R}} - i\langle if, g \rangle_{\mathbb{R}} = \big(\langle f, g \rangle_{\mathbb{R}} + i\langle if, g \rangle_{\mathbb{R}}\big)^* = \langle f, g \rangle^*,
\end{aligned}$$

where the third equation holds according to equation (17.12). To see the latter, it is enough to show that

$$\langle f, ig \rangle = i\langle f, g \rangle.$$

Indeed, we have

$$\begin{aligned}
\langle f, ig \rangle &= \langle f, ig \rangle_{\mathbb{R}} + i\langle if, ig \rangle_{\mathbb{R}} = \langle if, i^2 g \rangle_{\mathbb{R}} + i\langle i^2 f, i^2 g \rangle_{\mathbb{R}} \\
&= \langle if, -g \rangle_{\mathbb{R}} + i\langle -f, -g \rangle_{\mathbb{R}} = -\langle if, g \rangle_{\mathbb{R}} + i\langle f, g \rangle_{\mathbb{R}} \\
&= i\big(i\langle if, g \rangle_{\mathbb{R}} + \langle f, g \rangle_{\mathbb{R}}\big) = i\langle f, g \rangle,
\end{aligned}$$

where the second equation holds according to equation (17.12).

Solution of Problem 17.5. According to Minkowski's inequality (Clause (c) of Definition 17.2) we have

$$\|f\| = \|(f - g) + g\| \le \|f - g\| + \|g\|,$$

and so

$$\|f\| - \|g\| \le \|f - g\|.$$

Similarly,

$$\|g\| - \|f\| \le \|g - f\| = \|f - g\|.$$

Putting the last two inequalities together, inequality (17.14) follows.

Solution of Problem 17.6. According to equation (17.14), we have

$$|\|f_n\| - \|f\|| \le \|f_n - f\| \to 0,$$

which is what we wanted to show.

Solution of Problem 20.1. We have

$$
\begin{aligned}
(31.7) \qquad |f + g|^2 - |f - g|^2 &= (f + g)^*(f + g) - (f - g)^*(f - g) \\
&= (f^* + g^*)(f + g) - (f^* - g^*)(f - g) \\
&= (f^* f + f^* g + g^* f + g^* g) - (f^* f - f^* g - g^* f + g^* g) = 2f^* g + 2g^* f.
\end{aligned}
$$

Using this with if replacing f we obtain

$$|if + g|^2 - |if - g|^2 = -2if^* g + 2g^* if.$$

Multiplying the second equation by i and adding the resulting equations, we obtain equation (20.9).

Solution of Problem 20.2. Similarly to equation (31.7), we have

$$
\begin{aligned}
\|f + g\|^2 - \|f - g\|^2 &= \langle f + g, f + g \rangle - \langle f - g, f - g \rangle \\
&= \big(\langle f, f \rangle + \langle f, g \rangle + \langle g, f \rangle + \langle g, g \rangle \big) - \big(\langle f, f \rangle - \langle f, g \rangle - \langle g, f \rangle + \langle g, g \rangle \big) \\
&= 2\langle f, g \rangle + 2\langle g, f \rangle.
\end{aligned}
$$

Using this with if replacing f, we obtain

$$\|if + g\|^2 - \|if - g\|^2 = 2\langle if, g \rangle + 2\langle g, if \rangle = -2i\langle f, g \rangle + 2i\langle g, f \rangle.$$

Multiplying the second equation by i and adding the resulting equations, we obtain equation (20.12).

Solution of Problem 20.3. The necessity of equation (20.13) for the norm $\| \cdot \|$ to be a norm induced by a real- or complex-valued inner product on a vector space V over $\mathbb{R}$ or $\mathbb{C}$ can be easily established. Indeed, assuming that

$$\|f\|^2 = \langle f, f \rangle \qquad \text{for all} \quad f \in V$$

for a real- or complex-valued inner product, for all $f, g \in V$ we have

$$
\begin{aligned}
\|f + g\|^2 + \|f - g\|^2 &= \langle f + g, g + f \rangle - \langle f - g, g - f \rangle \\
&= \big(\langle f, f \rangle + \langle f, g \rangle + \langle g, f \rangle + \langle g, g \rangle \big) + \big(\langle f, f \rangle - \langle f, g \rangle - \langle g, f \rangle + \langle g, g \rangle \big) \\
&= 2\langle f, f \rangle + 2\langle g, g \rangle = 2\|f\|^2 + 2\|g\|^2.
\end{aligned}
$$

133

To show that equation (20.13) is sufficient for $\|.\|$ to be induced by an inner product in case V is a normed vector space over $\mathbb{R}$, define a putative inner product as

$$(31.8) \qquad \langle f, g \rangle \stackrel{def}{=} \frac{1}{4}\left(\|f+g\|^2 - \|f-g\|^2\right) \qquad \text{for all} \quad f, g \in V.$$

It is easy to see that we then have $\langle f, f \rangle = \|f\|^2$, so if $\langle \cdot, \cdot \rangle$ is an inner product, then it induces the norm $\|\cdot\|$. We need to show that $\langle \cdot, \cdot \rangle$ is indeed an inner product, i.e., that it satisfies the clauses in Definition 17.1. This is clear for Clauses (a) and (b), the latter since $\langle \cdot, \cdot \rangle$ is real valued and symmetric. Next, we will establish Clause (d).

To this end, we will first show that

$$(31.9) \qquad \langle f, g \rangle + \langle f, h \rangle = \frac{1}{2}\langle 2f, g + h \rangle.$$

We have

$$\langle f, g \rangle + \langle f, h \rangle = \frac{1}{4}\left(\|f+g\|^2 - \|f-g\|^2\right) + \frac{1}{4}\left(\|f+h\|^2 - \|f-h\|^2\right)$$

$$= \frac{1}{4}\left(\|f+g\|^2 + \|f+h\|^2\right) - \frac{1}{4}\left(\|f-g\|^2 + \|f-h\|^2\right)$$

Using equation (20.13), the right-hand side becomes

$$\frac{1}{8}\left(\|(f+g)+(f+h)\|^2 + \|(f+g)-(f+h)\|^2\right)$$

$$- \frac{1}{8}\left(\|(f-g)+(f-h)\|^2 + \|(f-g)-(f-h)\|^2\right)$$

$$= \frac{1}{8}\left(\|2f+g+h\|^2 + \|g-h\|^2\right) - \frac{1}{8}\left(\|2f-g-h\|^2 + \|-(g-h)\|^2\right)$$

$$= \frac{1}{8}\left(\|2f+(g+h)\|^2 - \|2f-(g-h)\|^2\right) = \frac{1}{2}\langle 2f, g+h \rangle,$$

where the last equation follows from equation (31.8). This verifies equation (31.9).

Now, it is easy to see from equation (31.8) that $\langle f, 0 \rangle = 0$. Hence

$$\langle f, g \rangle = \langle f, g \rangle + \langle f, 0 \rangle = \frac{1}{2}\langle 2f, g \rangle,$$

where the last equality holds by equation (31.9). Putting this together with equation (31.9), we obtain

$$(31.10) \qquad \langle f, g \rangle + \langle f, h \rangle = \langle f, g + h \rangle,$$

which establishes Clause (d) in Definition 17.1.

By repeated addition, equation (31.10) implies that

$$(31.11) \qquad \langle f, \alpha g \rangle = \alpha \langle f, g \rangle.$$

for every positive integer α. As

$$\langle f, g \rangle + \langle f, -g \rangle = \langle f, 0 \rangle = 0,$$

134

this equation also follows for all negative integers n. Hence we can conclude that

$$\langle f, g \rangle = \frac{1}{\alpha} \left\langle f, \frac{1}{\alpha} g \right\rangle$$

for all nonzero integers α by replacing g with $(1/\alpha)g$ in equation (31.11). Therefore, we can conclude equation (31.11) for all rational α.

To verify equation (31.11) for a given irrational α, let α_n be a sequence of rationals such that $\alpha_n \to \alpha$. Then we have

$$\lim_{n \to \infty} \|(\alpha_n - \alpha)g\| = \lim_{n \to \infty} |\alpha_n - \alpha| \, \|g\| = 0.$$

Hence, using equation (31.11) for rational α_n replacing α, we have

$$\alpha \langle f, g \rangle = \lim_{n \to \infty} \alpha_n \langle f, g \rangle = \lim_{n \to \infty} \langle f, \alpha_n g \rangle = \lim_{n \to \infty} \frac{1}{4} \left(\|f + \alpha_n g\|^2 - \|f - \alpha_n g\|^2 \right)$$

$$= \frac{1}{4} \left(\|f + \alpha g\|^2 - \|f - \alpha g\|^2 \right) = \langle f, \alpha g \rangle;$$

here the third and fifth equations hold according to equation (31.8), and the fourth equation holds according to equation (17.15). Thus equation (31.11) follows also for irrational α. This establishes Clause (c) of Definition 17.1, completing the proof that $\langle \cdot, \cdot \rangle$ is a real-valued inner product on V over $\mathbb{R}$.

Solution of Problem 20.4. The necessity of equation (20.13) was already established in the solution of Problem 20.3. Assuming that this equation is satisfied and considering V as a normed vector space over $\mathbb{R}$, it also follows from (the solution of) Problem 20.3. that there is an inner product $\langle \cdot, \cdot \rangle_{\mathbb{R}}$ satisfying equation (31.8). That is,

$$\langle f, g \rangle_{\mathbb{R}} = \frac{1}{4} \left(\|f + g\|^2 - \|f - g\|^2 \right) \qquad \text{for all} \quad f, g \in V.$$

As $\|if\| = |i| \|f\| = \|f\|$ for all $f \in V$, this equation implies that

$$\langle f, g \rangle_{\mathbb{R}} = \langle if, ig \rangle_{\mathbb{R}} \qquad \text{for all} \quad f, g \in V.$$

Hence $\langle \cdot, \cdot \rangle_{\mathbb{R}}$ can be extended to a complex-valued inner product according to Problem 17.4.

Solution of Problem 20.5 Using equation (20.1), we have

$$\hat{f}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(y) e^{-ixy} \, dy, = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\alpha^2 y^2} e^{-ixy} \, dy$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\alpha^2 y^2 - ixy} \, dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left(\alpha y - ix/(2\alpha)\right)^2 - x^2/(2\alpha)^2} \, dy$$

$$= e^{-x^2/(2\alpha)^2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left(\alpha y - ix/(2\alpha)\right)^2} \, dy = \frac{1}{\alpha} e^{-x^2/(2\alpha)^2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2} \, dt;$$

one can think of the last step here as using the substitution $t = \alpha y - ix/(2\alpha)$ where y and t are the variables and x and α are parameters, but for a rigorous justification one needs to use line integrals in the complex plane. The integral on the right-hand side is $\sqrt{\pi}$ according to equation (2.3); hence

$$\hat{f}(x) = \frac{1}{\alpha \sqrt{2}} e^{-x^2/(2\alpha)^2}.$$

135

References

[1] Selin Aviyente. Two-dimensional wavelets.
`http://www.egr.msu.edu/~aviyente/ece802tdw.ppt`, 2010.

[2] William Bell. A note on overdifferencing and the equivalence of seasonal time series models with monthly means and models with $(0, 1, 1)_{12}$ seasonal parts when $\theta = 1$. *Journal of Business & Economic Statistics*, 5(3):383–387, 1987. JSTOR stable url:
`http://www.jstor.org/stable/1391613`.

[3] Tim Bollerslev. Glossary to ARCH (GARCH).
`ftp://ftp.econ.au.dk/creates/rp/08/rp08_49.pdf`, September 2008.

[4] George E. P. Box, Gwilym M. Jenkins, and Gregory C. Reinsel. *Time Series Analysis. Forecasting and Control.* Prentice Hall, Englewood Cliffs, NJ 07632, third edition, 1994.

[5] Peter J. Brockwell and Richard A. Davis. *Time Series: Theory and Methods.* Springer-Verlag, New York, 1987.

[6] Chris Chatfield. *The Analysis Time Series. An Introduction.* Texts in Statistical Science. Chapman & Hall/CRC Press, New York, sixth edition, 1983.

[7] Liu Chun-Lin. A tutorial of the wavelet transform.
`http://disp.ee.ntu.edu.tw/tutorial/WaveletTutorial.pdf`, February 23, 2010.

[8] Barry Cipra. Engineers look to Kalman filtering for guidance.
`https://www.cs.unc.edu/~welch/kalman/siam_cipra.html`, August 1993.

[9] John J. Cochrane. A brief parable of over-differening.
http://faculty.chicagobooth.edu/john.cochrane/..., July 27, 2012. Non-clicable full url:
`http://faculty.chicagobooth.edu/john.cochrane/research/papers/overdifferencing.pdf`.

[10] John J. Cochrane. The grumpy economist. http://johnhcochrane.blogspot.com/..., January 26, 2012. Non-clicable full url: `http://johnhcochrane.blogspot.com/2012/01/brief-parable-of-over-differencing.html`.

[11] Paul S. P. Cowpertwait and Andrew V. Metcalfe. *Introductory Time Series with R.* Springer, New York, 2009.

[12] Jonathan D. Cryer and Kung-Sik Chan. *Time Series Analysis With Applications in R.* Springer Texts in Statistics. Springer, New York, second edition, 2010.

[13] Ingrid Daubechies. Orthonormal bases of compactly supported wavelets. *Comm. Pure & Appl. Math.*, 41 (7):909–996, 1988.
Posted at `https://services.math.duke.edu/~ingrid/publications/index.html`.

[14] Ingrid Daubechies. *Ten Lectures on Wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics.* Society for Industrial and Applied Mathematics, Philadelphia, 1992.

[15] Bradley Efron. Computers and the theory of statistics: Thinking the unthinkable. *SIAM Review*, 21(4):460–480, 1979. JSTOR stable url: `http://www.jstor.org/stable/2030104`.

[16] Robert R. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrics*, 50(4):987–1007, 1982. Published by The Econometric Society. Stable URL: `http://www.jstor.org/stable/1912773`.

[17] Ramsey Faragher. Understanding the basis of the kalman filter via a simple and intuitive derivation. http://www.cl.cam.ac.uk/ rmf25/papers/..., accessed on August 17, 2017. Non-clickable full url: `https://www.cl.cam.ac.uk/~rmf25/papers/Understanding%20the%20Basis%20of%20the%20Kalman%20Filter.pdf`.

[18] G. H. Hardy. *A Mathematician's Apology*. Cambridge University Press, Cambridge, 1969.

[19] Rob Hyndman, Anne B. Koehler, J. Keith Ord, and Ralph D. Snyder. *Forecasting with Exponential Smoothing: The State Space Approach*. Springer Series in Statistics. Springer, New York, 2008th edition edition, 2008.

[20] Claudia Kirch and Dimitris N. Politis. Tft-bootstrap: Resampling time series in the frequency domain to obtain replicats in the time domain. *The Annals of Statistics*, 39(3):1427–1470, 2011. `https://arxiv.org/pdf/1211.4732.pdf`.

[21] Jens-Peter Kreiss and Soumendra Nath Lahiri. Bootstrap methods for time series. *Time Series Analysis: Methods and Applications*, 30, 2012. `https://www.kevinsheppard.com/images/0/0a/Kreiss_and_lahiri.pdf`.

[22] Philippe Masset. Analysis of financial time-series using fourier and wavelet methods. *Social Science Research Network*, August 2008. `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1289420`.

[23] Attila Máté. Introduction to numerical analysis with C programs. `http://www.sci.brooklyn.cuny.edu/~mate/nml/numanal.pdf`, August 2013.

[24] Attila Máté. The cyclic decomposition theorem. `http://www.sci.brooklyn.cuny.edu/~mate/misc/cyclic_decomposition.pdf`, December 2014.

[25] Attila Máté. The natural exponential function. `http://www.sci.brooklyn.cuny.edu/~mate/misc/exp_x.pdf`, September 2015.

[26] The greatest math mistake. `http://www.sci.brooklyn.cuny.edu/~mate/misc/greatest_mistake.html`, 1999.

[27] Dimitris N. Politis. The impact of bootstrap methods on time series analysis. *Statistical Science*, 18(2):219–230, 2003. `http://projecteuclid.org/euclid.ss/1063994977`.

[28] Mikko Ranta. *Wavelet Multiresolution Analysis of Financial Time Series*. PhD thesis, University of Vaasa, Finland, 2010. `http://www.uva.fi/materiaali/pdf/isbn_978-952-476-303-5.pdf`.

[29] Abraham Robinson. *Non-Standard Analysis*. North Holland, New York, revised edition, 1974.

[30] Hans Schneider and George Philip Barker. *Matrices and Linear Algebra, 2nd ed.* Dover Publications, New York, 1973.

[31] Reginald Smith. Is high-frequency trading inducing changes in market microstructure and dynamics? *Social Science Research Network*, September 21, 2010.
`https://arxiv.org/pdf/1006.5490.pdf` with clickable internal links, and, without links: `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1632077`.

[32] Chong Tan. Financial time series forecasting using improved wavelet neural network. Master's thesis, Aarhus University, Aarhus, Denmark, 2009.
`http://www.cs.au.dk/~cstorm/students/Chong_Jul2009.pdf`.

[33] Ruey S. Tsay. Business 41910: Time series analysis for forecasting and model building, Autumn Quarter of 2008.
`http://faculty.chicagobooth.edu/ruey.tsay/teaching/uts/`. Course given at the University of Chicago Booth School of Business.

[34] Gary Welch and Gary Bishop. An introduction to the Kalman filter.
`http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf`, updated July 24, 2006.

[35] Wikipedia. Squirrel-cage rotor.
`https://en.wikipedia.org/wiki/Squirrel-cage_rotor`, accessed August 13, 2016.

[36] Wikipedia. Kalman filter.
`https://en.wikipedia.org/wiki/Kalman_filter`, accessed August 15, 2016.

[37] Wikipedia. Invariant subspace problem.
`https://en.wikipedia.org/wiki/Invariant_subspace_problem`, accessed August 17, 2016.

[38] Wikipedia. Matrix calculus.
`https://en.wikipedia.org/wiki/Matrix_calculus`, accessed August 17, 2016.

[39] Wikipedia. Criticism of non-standard analysis.
`https://en.wikipedia.org/wiki/Criticism_of_non-standard_analysis`, accessed August 18, 2016.

[40] Wikipedia. Non-standard analysis.
`https://en.wikipedia.org/wiki/Non-standard_analysis`, accessed August 18, 2016.

[41] Wikipedia. Levinson recursion.
`https://en.wikipedia.org/wiki/Levinson_recursion`, accessed July 26, 2016.